The Dissertation Committee for Mariana Mira Vasconcellos certifies that this is the approved version of the following dissertation:

Diversification and Evolution of Treefrogs in the Cerrado Savanna of South America: From Population Structure to Biogeographic Patterns

Committee:

David Cannatella, Supervisor

Claus Wilke, Co-Supervisor

Daniel Bolnick

Beryl Simpson

Kenneth Young

Michael Ryan

Diversification and Evolution of Treefrogs in the Cerrado Savanna of South America: From Population Structure to Biogeographic Patterns

by

Mariana Mira Vasconcellos, B.S.; M.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

The University of Texas at Austin

May 2016

Dedication

I dedicate this work to my family in Brazil, especially my parents Célia and Amilcar Vasconcellos, my sister Beatriz Vasconcellos and my aunt Léia Mira for their unconditional support, love, and understanding throughout my PhD.

Acknowledgements

A PhD is quite a journey! But luckily, I could count on the help of many people along the way, who contributed substantially during different steps towards the completion of this dissertation. First, this work was only made possible due to a long-term collaboration with Guarino Colli at UnB, who helped with all the logistics of fieldwork in Brazil, including collection permits, transportation, and support from his lab in Brazil. I would like to thank my advisor David Cannatella, my co-advisor Claus Wilke, and my committee members Mike Ryan, Dan Bolnick, Beryl Simpson, and Ken Young, for their guidance, support and invaluable advice, which greatly contributed to improve this work.

I thank CAPES (Brasil) and Fulbright (US) for providing financial support during the first four years of my PhD, which also allowed me to receive support from the Graduate Dean's Prestigious Fellowship at UT. My dissertation work has also received support from NSF (Doctoral Dissertation Improvement Grant DEB 1311517), FAP-DF (Brazil), CNPq (Brazil), and the Zoology Scholarship Endowment for Excellence from EEB-UT.

Several people have assisted me during key steps of data collection in the field, in the lab, and during data analysis. I am grateful to Mario van Gastel, Ísis Arantes, Tamara Vieira, Almir de Paula, Diego Santana, and Eric Barbalho for their invaluable assistance in the field. I am also grateful to Beatriz Vasconcellos and Ísis Arantes for their assistance in the morphological data collection at CHUNB. Patricia Salerno and Monica Guerra also greatly contributed with call data analysis. Lastly, I would like to thank Jesse Weber and Edgardo Ortiz for their generous help on the genomics part of my dissertation, and the support I received from GSAF and CCBB staff for genomic data collection and

analysis, especially Eric Ortego, Dennis Wylie and James Derry from CCBB, and Jessica Ponar and Scott Hunicke-Smith from GSAF. I was extremely lucky to have these incredibly talented people around to help me during genomic data processing and programming.

My lab mates and friends have also constantly helped me to improve my ideas through constructive feedback. Among them, I would like to give special thanks to Carlos Guarnizo and Brian Caudle for early lab work support, Patricia Salerno, Becca Tarvin, Monica Guerra, Jeanine McLean, Sofía Rodrígues, and Tom Devitt for helpful discussions about my work and manuscripts. Also, several other friends have made my time in Austin unforgettable and I will name some of them here: Kelsey Jiang, Yui Matsumoto, George Livingston, Laura Dugan, Vicky Huang, Simone and Thiago Carvalho, Lisiane and Joel Shuler, Roberta Veiga, Marco Aurélio, Deise Golçalves, Décio Thadeu, Vanessa Rivera, Oscar Vargas, among many others. My friends in Brazil have also always supported me from distance. Furthermore, I am grateful to my friends from Universidade de Brasília, some of whom are ongoing research collaborators: Fabrícius Domingos, Daniel Mesquita, Fernanda Werneck, Davi Pantoja, Laís Veludo, Marcela Brasil, Marcella Santos, Renan Bosque, Ísis Arantes, Gabriel Horta, among others.

Last but not least, I thank my family in Brazil for their love and support in my decision to leave Brazil and pursue a PhD in the US. The encouragement of my mom Célia Mira Vasconcellos, my dad Amilcar Vasconcellos, and my sister Beatriz Vasconcellos has always been crucial to me. My aunt Léia Mira also played an important role, cheering for me all the way. I will always carry her in my heart! Finally, the support and love I have received from my fiancée Mario van Gastel during all these years has been critical to my success. Thank you all!

Diversification and Evolution of Treefrogs in the Cerrado Savanna of South America: From Population Structure to Biogeographic Patterns

Mariana Mira Vasconcellos, PhD.

The University of Texas at Austin, 2016

Supervisor: David C. Cannatella

Co-Supervisor: Claus O. Wilke

Historical and ecological factors underlying population structure and speciation of organisms are fundamental to uncover diversification mechanisms that lead to biogeographic patterns. The main goal of this dissertation is to determine the relative roles of geography and climate promoting diversification in a diverse group of treefrogs in the Cerrado savanna of South America, a hotspot of biodiversity. This dissertation approaches diversification at two different time-scales: from a macro-temporal scale of factors that promote species diversification in the Cerrado region, and from a micro-temporal scale of population divergence and recent cryptic speciation. Three main research goals were developed in separate chapters: (1) Phylogenetic and biogeographic analyses of the species in the *Hypsiboas pulchellus* group were used to quantify dispersal events between the Cerrado and other dry/open or humid/forested neighboring habitats, and to evaluate how this dynamic historical interchange affected diversification. The evolutionary relationships among the species in this group showed a pattern of multiple recent origins of the endemic Cerrado lineages, indicating recent colonization events.

These analyses indicate an interesting pattern of recurrent dispersal among open and forest ecoregions accounting for the accelerated diversification in this group. (2) Species delimitation methods using molecular, morphological and mating call characters were used to study the recent speciation of a small clade of Cerrado endemics. This integrative approach allowed for a better resolution to detect species boundaries and the discovery of cryptic species in spite of conserved morphology and mating calls. (3) Population structure in a widespread Cerrado species was investigated using historical species distribution modeling and next generation sequencing data to evaluate the role of recent climatic fluctuations on population differentiation. Divergence among populations seems strongly affected by Pleistocene climatic instability, a pattern we call isolation-by-instability. In conclusion, this study highlights how diversification and biogeographic patterns in the Cerrado savanna can be affected by its historical dynamic climate and landscape.

Table of Contents

List of Tables

List of Figures

INTRODUCTION

Understanding the historical and ecological factors underlying the genetic structure of organisms is fundamental to identify the mechanisms leading to diversification and biogeographic patterns. It can also shed light on important drivers of the incipient process of recent diversification of species and populations (Graham *et al.* 2004). Recent advances in coalescent-based analyses incorporating the heterogeneity of gene trees coalescent times have allowed a more robust estimation of the evolutionary history of species and populations (Liu *et al.* 2008). Additionally, integrating present and past distribution models has represented exciting advances in testing diversification hypotheses. However, the new approaches of using whole genomic data instead of a just few genetic markers have been delayed in phylogeographic and speciation studies (McCormack *et al.* 2013), in part due to the uncertainty of methods recovering homologous loci across samples.

Herein, I use recently developed sequencing technologies to uncover diversification hypotheses in a widespread treefrog species in a Neotropical savanna. I also make use of predictive distribution models associated with phylogenetics and phylogeography to study the historical and ecological factors promoting diversification in a treefrog lineage in South American savannas at two different time scales: on a macro-temporal scale of a deep species phylogeny, and on a micro-temporal scale of population divergence. My main dissertation goals are thee-fold: 1) Examine the roles of dispersal that have shaped the historical biogeography of a treefrog lineage across South American ecoregions with a special focus on savanna endemics; 2) Determine species limits in a lineage of Cerrado endemics understanding how morphological and behavioral divergence contributes to the speciation process; and, 3) Determine the historical and

geographical factors promoting population divergence in South American savannas using Next-Gen sequencing to evaluate the genetic structure of a treefrog and the effects of recent Pleistocene climatic fluctuations and the role of geographic barriers to gene flow.

The Neotropical region encompasses a large fraction of global biodiversity, however, knowledge about the mechanisms and processes involved in the origin and diversification of the Neotropical biota is still in its infancy. In a global survey, Beheregaray (2008) identified a huge gap in phylogeographic studies in South America, which is the least-studied continent (excluding Antarctica) for phylogeographic studies, while ironically, harboring the greatest biodiversity on Earth. Although some significant studies have emerged, the long-standing controversy of the impact of recent climate fluctuations on the genetic diversification of South American organisms is still a recurrent debate (Carnaval & Moritz 2008; Werneck *et al.* 2012a). The extent to which glaciations have impacted South America remains unclear. Even though glaciers might not have covered most of South America, cycles of contraction and expansion of forests and savannas during glacial and interglacial periods are thought to have occurred (Van der Hammen 1974) and greatly affected the vegetation structure and distribution of organisms (Bonaccorso *et al.* 2006; Saia *et al.* 2008). Although some studies have found evidence for climatic fluctuations affecting population structure in refuges of historical stable climate in the Atlantic rainforest (Carnaval & Bates 2007; Carnaval *et al.* 2009; Martins 2011), others have failed to support this hypothesis in other regions (Thomé *et al.* 2010; Werneck *et al.* 2012a). However, most of these studies have focused on the Atlantic rainforest, with few studies addressing how climatic fluctuations might have impacted open vegetation biomes of South America.

The largest and most species-rich savanna on Earth, the Cerrado savanna of South America, is one of the 25 global hotspots of biodiversity highly threatened by

deforestation (Myers *et al.* 2000; Klink & Machado 2005). Evolutionary or ecological processes responsible for this great diversification are not well studied and generally based on biogeographic methods not considering the evolutionary history of the organisms (Silva & Bates 2002; Colli 2005). This savanna is part of an open dry diagonal that separates the Amazon forest in the northwest and the Atlantic forest in the southeast. It is composed of a mosaic of open to closed vegetation ranging from grasslands to gallery forest along rivers, and is remarkable for the high endemicity level reported for several organisms including: plants (44%; Klink & Machado 2005), frogs (53%; Valdujo *et al.* 2012), and lizards (39%; Nogueira *et al.* 2011).

Two main hypotheses have been proposed to explain the high observed unique diversity in this savanna: (1) the old origin of the open Cerrado vegetation predating the surrounding rainforests enabling enough time for in-situ diversification (Silva & Bates 2002); (2) the more recent diversification of Cerrado endemics through adaptive shifts from surrounding biomes coinciding with the predominance of flammable C4 grasses and expansion of the savanna biome worldwide (Simon *et al.* 2009). Little evidence exists for either hypothesis because of the lack of studies dating divergence of Cerrado endemics. Moreover, the relationship of the Cerrado fauna and flora with other open or forest biomes is not well understood.

Few recent studies focusing on diversification processes of the Cerrado ecoregion at either deep or shallow time-scales have yielded mixed results about the importance of factors driving diversification in this habitat. While some have highlighted the importance of Miocene-Pliocene geological events shaping the deep relationships among many taxa in this habitat (Maciel *et al.* 2010; Werneck *et al.* 2012a), others have showed the relatively recent radiation of many endemic taxa emphasizing the importance of recent events such as climatic fluctuations on the Pleistocene or recent adaptation to fire

and drought promoting ecological radiation (Simon *et al.* 2009; Prado *et al.* 2012). However, no study has used genomic data to evaluate gene flow and test alternative ecological and geographical factors structuring populations in this habitat. The use of high-throughput sequencing can improve population genetic estimates allowing a better inference of population structure that would be otherwise biased by using just a few genetic markers.

Treefrogs can be great models to study the impact of environment and geography on population and species divergence. Because of their low vagility, high phylopatry to natal sites, and because their populations are often highly structured in breeding assemblages, they can be useful indicators of environmental variation in time and space (Beebee 1995). Therefore, populations can retain high-resolution genetic signatures of historical events that helped generating current species distributions (Zeisset & Beebee 2008). Hylid treefrogs also consist of an eminent example of Neotropical radiation that has inspired several studies about the evolutionary and ecological causes underlying tropical community assembly and latitudinal gradients of diversity (Wiens *et al.* 2006; Moen *et al.* 2009). Among its most speciose groups, the *Hypsiboas pulchellus* group (Wiens *et al.* 2010) are frequently associated with open habitats in South America (Faivovich *et al.* 2004), moreover the causes of this high diversity in barrier free environments like savannas is urged to be addressed.

The *Hypsiboas pulchellus* group currently comprises 38 species, including a group of 9 species restricted to the Brazilian Shield previously assigned to the *Hypsiboas polytaenius* group, which is thought to be monophyletic based only on their morphological similarity (Cruz & Caramaschi 1998; Faivovich *et al.* 2005). I report on their evolutionary relationships including most species in this group, especially all the Cerrado endemics that have not been assessed before, and find an unexpected

relationship among them. The *Hypsiboas polytaenius* group comprises three clades within the *H. pulchellus* group, which is a result of several dispersal events across divergent habitats in South America. In the following chapters, I evaluate the main mechanisms involved in the diversification of this treefrog group in the Cerrado savanna of South America.

Chapter 1: Recurrent dispersal and diversification of treefrogs across distinctive biodiversity hotpots in sub-Amazonian South America

ABSTRACT

Although the Neotropical region encompasses a large fraction of global biodiversity, knowledge about the origin and diversification of its biota is still limited, particularly in dry and open areas such as savanna regions. Some general hypotheses have been proposed to describe species diversification patterns in diverse tropical regions of high endemicity, such as center of origin or museums/cradles of biodiversity, among others. We tested some of those diversification hypotheses exploring the biogeographic history of a South American treefrog lineage, the *Hypsiboas pulchellus* group, which is distributed in both open and forested ecoregions, including three biodiversity hotspots. This clade originated during the middle Miocene climatic transition, most likely in the combined area comprising the Cerrado savanna, the Central Andes, and the Araucaria Forest. We found no support that the ecoregion of richest diversity, the Araucaria Forest, acted as a center of origin. Diversification was not constant over time, but rather followed a logistic model of diversity-dependent cladogenesis with increased diversification coinciding with the origin of C_4 grasslands and expansion of savannas worldwide. Our ancestral range reconstruction analyses indicate repeated dispersal among dry/open and humid/forested ecoregions rather than long-term *in-situ* diversification in single areas. This highlights the dynamic historical interchange between contrasting habitats in South America.

INTRODUCTION

The Neotropical region encompasses about one third of the 25 biodiversity hotspots on Earth (Myers *et al.* 2000). Explanations for the high species diversity in this region have traditionally focused on rainforest habitats (Haffer 1969; Moritz *et al.* 2000; Hoorn *et al.* 2010; Wiens *et al.* 2011; Fine *et al.* 2014) with relatively little attention to other habitats such as savannas or dry tropical forest (Hughes *et al.* 2013). For example, the Cerrado hotspot of South America, one of the largest and most species-rich savannas on Earth (Myers *et al.* 2000; Oliveira & Marquis 2002), is largely understudied, and historical factors responsible for its great diversity are still not well understood. Only recently have dated phylogenies of endemic Cerrado taxa become available to permit inferences of evolutionary processes generating its unique biota (Simon *et al.* 2009; Hughes *et al.* 2013). The emerging pattern for plants is a recent assembly and diversification of the typical Cerrado vegetation promoted by adaptations to fire (Simon *et al.* 2009; Simon & Pennington 2012).

To understand biodiversity patterns in species-rich regions such as the Neotropics, we need to understand the processes responsible for generating and maintaining this diversity, a challenging endeavor given the large number of species and scant knowledge about them. Since regional biotas arise through a combination of macroevolutionary processes including speciation, extinction and dispersal (Ricklefs 1987), species richness patterns in the Neotropics should be evaluated based on clear predictions of those processes. This requires well-sampled, well-supported, and time-calibrated phylogenies. Even though their numbers are increasing, few estimates of species diversification rates exist for Neotropical clades, making generalizations difficult (Hughes *et al.* 2013). Still, some general predictions have been proposed to explain diversity patterns in tropical regions and we address these below.

Traditionally, tropical regions have been claimed to act either as diversity accumulators with low extinction rates (the museum hypothesis) or diversity pumps with high rates of speciation (the cradle hypothesis) (Jablonski 1993). However, these hypotheses are not necessarily mutually exclusive. Increasing evidence suggests that tropical regions may act as both cradles and museums of biodiversity (Jablonski *et al.* 2006; McKenna & Farrell 2006; Moreau & Bell 2013). In addition, studies showing concomitant high speciation rates and low extinction rates for diverse tropical taxa such as marine bivalves, birds, butterflies, amphibians and mammals provide strong support for both hypotheses (Goldberg *et al.* 2005; Ricklefs 2006; Condamine *et al.* 2012; Pyron & Wiens 2013; Rolland *et al.* 2014).

On a smaller regional scale, a similar framework has been used to describe taxonomically diverse regions with extremely high endemicity levels. These may either represent "centers of origin" with higher rates of *in-situ* origination of species (similar to the cradle hypothesis) or "centers of accumulation," where most taxa are mainly acquired through high levels of immigration from nearby regions associated with low levels of *in-situ* extinction (Mora *et al.* 2003; Goldberg *et al.* 2005). The term "center of origin" has also been used to simply designate the geographic region where a lineage originated (Bremer 1992). But, as pointed out by Goldberg *et al.* (2005), the region of first appearance of a lineage might not be the region in which most diversification occurred; therefore, this second definition may not be sufficient to describe the high richness and endemism levels in certain regions.

Other patterns not requiring differential diversification or dispersal rates among regions have also been proposed to explain the high diversity of the Neotropical region. These include the time-and-area hypothesis, which postulates that regions with tropical climates are older and cover more area than any other region (for a review see

Rosenzweig 1995; Mittelbach *et al.* 2007). The time component predicts that species richness is simply a function of the time that a lineage has been evolving in a given region and the area component predicts that larger areas have the capacity to hold more species. This idea dates back to Wallace (1878), but variations exist. For example, the "time-for-speciation effect" (Stephens & Wiens 2003) has been suggested as one of the main drivers of species diversity, especially for Neotropical frogs, in which species richness is highly correlated with the time since the region was colonized (Wiens *et al.* 2006; 2011; Hutter *et al.* 2013). It strongly relies on the idea that niche conservatism helps maintaining the latitudinal diversity gradient since species tend to retain their ancestral tropical ecological niches (Wiens & Donoghue 2004).

Hylid treefrogs are a good example of a Neotropical radiation, with several studies addressing the evolutionary causes of high local species assemblages in the tropics and the effects on latitudinal diversity patterns (Wiens *et al.* 2006; Moen *et al.* 2009; 2011). In this study, we focus on the *Hypsiboas pulchellus* group, a sub-clade of the True Gladiator Frogs, which is among the most species-rich groups of hylids in southern South America (Faivovich *et al.* 2005). This lineage occurs in several diverse habitats such as savannas, forests, grasslands and montane habitats, including three hotspots of biodiversity: the Atlantic Forest (including its southern extension, the Araucaria Forest), the Cerrado, and the tropical Andes (which includes the Northern and Central Andes) (Myers *et al.* 2000). The place and time of origin have not been inferred for this group, but a general forest origin has been suggested (Faivovich *et al.* 2004) based on present forest distributions of most of its species. This clade is also of particular interest to South American biogeography because of its central-south distribution, whereas most analyses have focused only on the northern part of the continent encompassing the Amazon Forest.

Herein, we investigate the origin and evolutionary history of the *Hypsiboas pulchellus* group using a time-calibrated phylogeny, and biogeographic and diversification analyses. Using ancestral range reconstruction we test different scenarios of dispersal and diversification in this group including a "center-of-origin" diversification, testing whether origin and most diversification events occurred in a single area further dispersing to other regions. We also infer the historical dispersal pattern of this group across ecoregions, based on the spatial configuration (distance and adjacency) among them. Additionally, we estimate the temporal pattern of diversification in this lineage (constant or variable) and determine whether different speciation and/or extinction models explain species accumulation, relating these results to the predictions of the museum and cradle hypotheses. Finally, since this treefrog lineage includes a substantial number of savanna endemics, we discuss the implications of our results in species assembly of the overlooked Cerrado savanna.

MATERIALS AND METHODS

Taxon sampling

We sampled 31 of the 38 recognized species (82%) for the *Hypsiboas pulchellus* group. Of the species we could not obtain samples: two are categorized as either critically endangered (*H. cymbalum*) or near-threatened (*H. alboniger*), three are data deficient (*H. secedens*, *H. freicanecae, H. palaestes*) according to the Red List of threatened species (IUCN 2014) , and two (*H. poaju*, *H. bandeirantes*) are very recently described species with narrow distributions and yet to be included in the IUCN list. Nonetheless, we present here the most complete phylogenetic analysis of the *H. pulchellus* group, with widespread geographic sampling for most species.

Molecular data were collected from 47 individuals of the *H. pulchellus* group (sensu Faivovich *et al.* 2005), including 11 of the 12 recognized species in the *H. polytaenius* subgroup (sensu Cruz & Caramaschi 1998). The species in the *H. polytaenius* subgroup, all of which are small species with dorsal stripes, have been identified as a clade within the *H. pulchellus* group, but only three species have been evaluated so far (Faivovich *et al.* 2004; 2005). We included samples from different populations, including the type locality whenever possible, to detect taxonomic problems and identify potential cryptic species. Published GenBank sequences were also included (Supporting Information). Tissue samples were obtained either through field sampling deposited in the Coleção Herpetológica da Universidade de Brasília (CHUNB) or as loans from collections and museums (see Acknowledgments).

DNA extraction, amplification, and sequencing

Genomic DNA was extracted from liver or muscle using the Viogene Blood and Tissue Genomic DNA Extraction Kit, following the manufacturer's protocol. An overlapping set of primers was used to amplify a region of approximately 2400 base pairs encompassing the mitochondrial genes 12S rRNA, valine tRNA, and 16S rRNA (hereafter referred to as 12S-16S), using standard PCR protocols and the following primers: MVZ59, tRNA-val, 12L1, 16SH, 12SM, 16SA 16SC, 16SD (Darst & Cannatella 2004). We also included two nuclear markers. Part of the rhodopsin gene (~850bp) including exon 1, an intron, and exon 2, was amplified in one fragment using the primers: Rhod1U (5'-aacggaacagaaggcccaaactt-3') and Rhod 1L (5'-gccaaagccatgatccaggtga-3'). Part of the recombination-activating gene (RAG1) (~450bp) was amplified using the primers R1-GFF and R1-GFR (Faivovich *et al.* 2005). The thermocycler conditions for 12S-16S included an initial denaturation at 94°C for 2 min, followed by 35 cycles of

denaturation at 94°C for 30 sec, annealing at 48°C for 30 sec, extension at 72ºC for 1 min, and a final extension phase of 7 min at 72°C. Conditions for the rhodopsin gene were similar to the above except that the annealing phase was 55°C for 45sec. The conditions for RAG1 were also similar to the above apart from the annealing phase, in which a touchdown protocol was performed with varying temperatures in the first ten cycles, starting at 52ºC and decreasing 2ºC at each two cycles until reaching 44ºC, then performing another 30 cycles at 40ºC. Amplified PCR products were then purified using the Viogene Gel Extraction Kit (Viogene BioTek Corporation). Sequencing of forward and reverse fragments was performed on an ABI 3100 PRISM sequencer at the DNA Sequencing Facility at The University of Texas at Austin. Editing and assembly of contiguous sequences for each individual were performed using Sequencher (Gene Codes Corporation). GenBank accession numbers and voucher information for new and previously published molecular sequences are provided in Supporting Information. Field and lab procedures were approved by the Institutional Animal Care and Use Committee of The University of Texas at Austin (IACUC, AUP-2010-00143).

Phylogenetic inference

DNA sequences were aligned using MUSCLE v. 3.8.31 (Edgar 2004) available online at the EMBL-EBI Web Services (http://www.ebi.ac.uk/Tools/msa/muscle/). Because some areas of the 12S-16S genes were difficult to align, we used SATé-II (Liu *et al.* 2012), which optimizes the alignment based on a co-estimated maximum likelihood phylogenetic tree. The preferred alignment is the one that produces the tree with the highest ML score using a stopping rule of 20 iterations after the last improvement. The final combined dataset of 12S-16S, rhodopsin, and RAG1 consisted of 3844 aligned base pairs and was 95% complete. Because the partitioning scheme may affect the accuracy of

phylogenetic reconstruction (Brown & Lemmon 2007), we employed PartitionFinder v. 1.1.1 (Lanfear *et al.* 2012), which performs a heuristic search for the best-fit partitioning scheme with simultaneous substitution model selection. The best scheme, using a greedy heuristic search algorithm of 56 possible models based on the Bayesian information criterion (BIC) with linked branch lengths, included four partitions: (1) the 12S-16S mitochondrial genes under the GTR+G+I substitution model, (2) the first and third codon positions of RAG1 and rhodopsin under the K80+G+I model, (3) the second codon position of RAG1 and rhodopsin under the F81 model, and (4) the rhodopsin intron region under the HKY+G model. For subsequent analyses these partitions were used.

Bayesian analyses were performed using MrBayes v3.2.1 (Ronquist *et al.* 2012). To account for substitution model uncertainty, we used MCMC "model jumping" across the model space by setting nst = mixed. Two independent runs of five million generations were sampled at every 1000 generations using four Markov chains, using a temperature of 0.02 and default priors for other parameters. We used Tracer v1.6 (Rambaut & Drummond 2013) to assess convergence and stationarity among runs and to determine appropriate burn-in values (found to be 20%). Maximum likelihood analyses were performed in RAxML v.8 (Stamatakis 2014) using the four partitions described above and the GTRGAMMA model for each partition following the author's recommendations. Node support was assessed using Bayesian posterior probabilities in MrBayes and non-parametric bootstrap values in RAxML (1000 replicates). All phylogenetic analyses were conducted on the Phylocluster at the Center for Computational Biology and Bioinformatics (CCBB) at The University of Texas at Austin.

Divergence time estimation

Reliable Tertiary fossil records are very scarce for anurans. The few fossil records for the family Hylidae consist mostly of few disarticulated broken bones (Holman 2003) with few diagnostic characters, which makes their use as calibration points questionable. Therefore, we calibrated our tree using divergence times estimated from deep-scale anuran phylogenies using deeper and more reliable fossil calibrations (Supporting Information Table S2). We then pruned our matrix to a single representative of each species and augmented our dataset using GenBank sequences of the same genes (12S-16S, rhodopsin, RAG1) for an additional 73 species of hylids to encompass the calibration points (Table S3). A shorter rhodopsin sequence containing only exon 1 (~350bp) was retained for this dataset, since the intron and exon 2 were not available for most species. We spread our sampling broadly so as to cover the entire family, using the Pyronand Wiens (2011) tree as a guide.

To prune the *Hypsiboas pulchellus* group dataset to a single representative per species, we used a *p*-distance threshold of 3% in the 16S gene as the criterion for retaining individuals in the same nominal species, because this level of divergence might roughly delimit cryptic species (Fouquet *et al.* 2007). MEGA 5.2.2 (Tamura *et al.* 2011) was used to calculate the uncorrected pairwise genetic distances among individuals (*p*-distance). Our goal was not to identify cryptic species per se, but to use an objective criterion for inclusion of recently diverged lineages in the diversification analyses.

Divergence times for the Hylidae dataset were estimated using BEAST v. 1.7.5 (Drummond *et al.* 2012) under the following scheme as determined by PartitionFinder: (1) 12S-16S under the GTR+G+I model, (2) first and second codon positions of RAG1 and rhodopsin under the HKY+G+I model, (3) third codon position of rhodopsin under the TrNef+G model, and (4) third codon position of RAG1 under the K80+G model. A

slightly different partition scheme was recovered likely due to differences between datasets in the length of the rhodopsin sequence and the number of taxa. We examined inferred divergence times from several studies (San Mauro *et al.* 2005; Smith *et al.* 2005; Roelants *et al.* 2007; Heinicke *et al.* 2009; Santos *et al.* 2009; Zhang *et al.* 2013). Based on these, we selected four well-supported nodes to which to apply secondary calibrations: (1) the most recent common ancestor (MRCA) of Hylini; (2) the MRCA of the Hylini+Lophiohylini; (3) the MRCA of Pelodryadinae+Phyllomedusinae; (4) the MRCA of Hylidae. We chose to use the divergence estimates of Heinicke *et al.* (2009) for two reasons: all four nodes were included, and the mean divergence times from the other studies fell within the 95% credible intervals (CI) from Heinicke *et al.* (2009) (Table S2).

We used a normal distribution for the prior around the mean divergence times reported in Heinicke *et al.* (2009) for the four calibration points. The standard deviation of the prior was set to a value that, when multiplied by two, would encompass the 95% CI of each calibration time. To ascertain the appropriate molecular clock model, we ran BEAST under the assumption of a strict molecular clock, a lognormal relaxed clock and an exponential relaxed clock using the birth-death process of speciation with uninformative default priors except for the calibration points. Using the Bayes factor (BF) criterion (Kass & Raftery 1995), the lognormal relaxed clock model was selected with strong support (BF >> 10). BEAST analyses were run for 30 million generations sampled at every 3000 generations. Convergence was determined using an ESS >200 for all parameters as assessed with Tracer v1.6. Twenty percent of the samples were discarded as burn-in. The maximum clade credibility tree was summarized using TreeAnnotator v1.7.5 (BEAST software package) and visualized using FigTree v1.4.2 (Rambaut 2014).

Ancestral range reconstruction

To infer the biogeographic history of the *Hypsiboas pulchellus* group, we used ancestral range reconstruction analyses. The geographic range of each species was determined from the literature and from collection localities from herpetological collections in Brazil. We defined six ecoregions based on Udvardy (1975) and Olson *et al.* (2001) characterizing each one as dry or humid. With the exception of the Araucaria Forest and the Pampas, these correspond to the ecoregions recognized by the World Wildlife Fund (wwf.panda.org/about_our_earth/ecoregions/ecoregion_list): (A) the Central Andean Yungas, composed of the eastern cordillera (Cordillera Oriental) of the central Andes of Peru and Bolivia (humid), (B) the Southern Andes, composed of the Andean dry Puna in southern Bolivia and Argentina (dry), (C) the Cerrado savanna of central Brazil (dry), (D) the Brazilian Atlantic Forest (humid), (E) the Araucaria Forest in southeastern Brazil, including Misiones province in Argentina (humid), (F) the Pampas grasslands of southern Brazil, Uruguay and Argentina, including the sierras of central Argentina (dry). Although most species were assigned to a single region, two species have ranges comprising two adjacent regions.

To reconstruct ancestral ranges, we used the Dispersal-Extinction-Cladogenesis (DEC) model (Ree & Smith 2008) and the DEC+J model (Matzke 2014) that incorporates a founder-event jump dispersal, using the BioGeoBEARS R package (Matzke 2013). Both methods estimate the maximum likelihood ancestral range at each node of the tree and also a global rate of dispersal and extinction among areas. We initially performed an unconstrained dispersal analysis with equal probability of dispersal among all areas regardless of distance and adjacency. This was used as our null model (M0) to test two additional biogeographic scenarios by varying the dispersal rate matrix between areas (Fig. 1.1).

First, in the M1 model, we tested the "center of origin" hypothesis, which posits that the Araucaria Forest, which has the highest species richness for the *H. pulchellus* group (12 species), is the center of origin for this clade, in other words, the place where it originated and mostly diversified. Dispersal from the Araucaria Forest to a different area was modeled as ten times more likely than between any other pair of areas, with ancestral ranges constrained to include this region. To do this, we manually modified the BioGeoBEARS script to exclude ancestral areas that did not include the Araucaria Forest as an ancestral state (E); see Supporting Information. Second, in the M2 model, we constrained the dispersal matrix using a stepping-stone model to reflect adjacency; the pairwise distance between areas was modeled in 1, 2 or 3 step increments, with dispersal probability between areas decreased by half at each step. Ancestral ranges consisting of more than one area were also constrained to include spatially adjacent areas. In the M2 model, any area may be the source of species dispersal to other areas.

For all analyses, ancestral ranges were constrained to include no more than three areas, given that no extant species occupies more than two areas. The global likelihood of each model was then compared to the likelihood of the null model, with differences greater than two log-likelihood units considered significant (Edwards 1992; Ree & Smith 2008).

Diversification rates

Using a chronogram including only the *Hypsiboas pulchellus* clade (33 species), we evaluated several diversification models that address temporal variation in diversification rates using R 3.1.1 (R Development Core Team 2015). First, we generated a lineage-through-time (LTT) plot using the ape package (Paradis *et al.* 2004) to visually determine departure from constant diversification. Then, using the phytools package

(Revell 2011) we plotted the estimated LTT against the 95% confidence interval of the branching times across 1000 simulated birth-death trees scaled to the same number of terminals as the *Hypsiboas pulchellus* chronogram. To test the null hypothesis of constant diversification rates, we used the MCCR gamma statistics test (Pybus & Harvey 2000) against a null distribution of constant rate accounting for incomplete sampling in the laser package (Rabosky 2006). We then compared the fit of constant diversification models (Yule pure birth and birth-death models) to several variable-rate models (two-rate and three-rate Yule models, and both exponential [DDX] and logistic [DDL] diversity-dependent models). We also assessed if variable speciation and/or extinction could explain the temporal variation in diversification rates in this group, fitting models in which speciation and extinction rates vary independently of each other. For that, we fit models with constant speciation rate and declining extinction rate (EXVAR), declining speciation rate and constant extinction rate (SPVAR), and speciation and extinction rates that both varied through time (BOTHVAR). We evaluated all diversification models using the Akaike Information Criterion (AIC) through a standard model selection procedure in which best-fit models were selected based on lower AIC scores and greater AIC weight (Burnham & Anderson 2002).

The diversification pattern generated by lineage-through-time plots is usually assumed to mainly result from variable speciation rates (Quental & Marshall 2010). For example, diversity-dependent models generally assume no extinction since high levels of extinction would not allow the detection of such pattern (Rabosky & Lovette 2009). Despite this, we explicitly quantified the heterogeneity in speciation and extinction rates, also testing for possible among-lineage rate variation using BAMM (Rabosky 2014) and its companion R package BAMMtools (Rabosky *et al.* 2014). After detecting for a diversity-dependent model in the diversification of the *H. pulchellus* group (see Results),

we used models that explicitly test for the influence of extinction in the package DDD (Etienne *et al.* 2011). We fit models incorporating diversity-dependence as linearly dependent on extinction rate (DDL+E), speciation rate holding extinction rate constant (DDL–E), and both speciation rate and extinction rate (DDL+SE).

RESULTS

Phylogeny and divergence time estimation

Maximum likelihood and Bayesian phylogenetic inference recovered the same topology (Fig. 1.2). The species in the *Hypsiboas pulchellus* group form a strongly supported clade with the species in the *H. faber* group recovered as a sister clade with high support. Interestingly, our analyses reveal that the small, striped treefrogs which have been previously assigned to the *H. polytaenius* group (sensu Cruz & Caramaschi 1998) consist of a polyphyletic group, with three distinct and strongly supported lineages interspersed within the *H. pulchellus* group. Support for most nodes was high. The few nodes lacking strong support (BPP < 0.9 and bootstrap < 70) were the node leading to *H. guentheri*, *H. marginatus*, and *H. bischoffii*; the node leading to *H. buriti*, *H. polytaenius* and *H. stenocephalus*; and the relationship among the sub-clades in clade B (Fig. 1.2), in which the low support can be explained by the uncertain placement of the clade formed by *H. aguilari*, *H. gladiator*, *H. balzani* and *H. melanopleura*, which was either part of clade B or the sister lineage of clades A and B (alternative trees not shown). In either case, species occurring in the Andes consisted of two non-monophyletic groups. *Hypsiboas ericae* from the Cerrado is the most basal species of the *H. pulchellus* group and the sister species of the other two large clades: clade A, comprising species from the Cerrado, the Atlantic Forest and the Araucaria Forest; and clade B, comprising species

from the Central Andes, the Southern Andes, the Araucaria Forest, the Atlantic Forest, the Pampas and the Cerrado.

The inclusion of several populations from most species revealed some taxonomic issues. Specimens identified as H. *phaeopleura* and *H. goianus* seem to represent the same species (Fig. 1.2E; see Discussion). *Hypsiboas polytaenius* (the most widespread among the striped treefrogs) appears to be a species complex consisting of at least two well-diverged lineages. Here, we follow the recent taxonomic changes of Köhler *et al.* (2010) in which *H. riojanus* is considered the senior synonym of *Hypsiboas andinus*, and several populations previously recognized as *H. balzani* are allocated to *Hypsiboas callipleura* (Fig. 1.2, specimen DLR41193).

Our chronogram recovered a root age of 61.8 MY (95% HPD = 49.5–73.9 MY) for the family Hylidae. The topology agrees with previous phylogenies (Faivovich *et al.* 2005; Wiens *et al.* 2010; Pyron & Wiens 2011) recovering the three subfamilies: Hylinae (57.6 MY), Pelodryadinae (38.0 MY), and Phyllomedusinae (29.1 MY) (Fig. S1). The age of the most recent common ancestor of the *Hypsiboas pulchellus* group is 15.4 MYA in the Miocene (95% HPD = 11.9–19.4 MY). The divergence times of the Cerrado lineages range from species younger than 5 MYA, such as *H. buriti* and *H. jaguariaivensis*, to older lineages, such as *H. ericae*, which is estimated to have diverged at 15.4 MYA (Fig. S2).

Ancestral range reconstruction and dispersal scenarios

The ancestral range of the *Hypsiboas pulchellus* group was most likely a widespread area comprising the Cerrado (C), the Central Andes (A) and the Araucaria Forest (E) (Fig. 1.3). We did not find support for the Araucaria Forest as the center of origin for this group, since M1 model performed significantly worse for both DEC (lnL

M1 = −74.82) and DEC+J (lnL M1 = −55.62) compared to the null model M0 (DEC lnL M0 = −65.74; DEC+J lnL M0 = −52.70). We also found that the spatial configuration of ecoregions had a great influence on dispersal of these treefrogs, since the M2 stepping-stone model (DEC lnL M2 = −61.07; DEC+J lnL M2 = −48.73) was strongly supported over the M0 model, which does not account for such spatial structure. Therefore, we evaluated range evolution of the *Hypsiboas pulchellus* clade under the M2 model.

DEC+J was favored over DEC under M0, M1 and M2 scenarios. However, because this model was developed to accommodate the founder-event speciation mode (J parameter) common in oceanic island systems, one might argue that it is perhaps unrealistic for comparing continental ecoregions. Nonetheless, we contrast the results of the DEC and DEC+J models.

The DEC model favored ancestral range reconstructions of two or three areas, especially in the nodes close to the root, while the DEC+J model favored single-area ancestors (Fig. 1.3). Under the DEC model, six dispersal events by range expansion were identified: four to the Atlantic Forest (D), one to the Pampas (F) and one to the Southern Andes (B). Following the categories of range evolution and notation proposed by Matzke (2014), we identified the following events under DEC: 8 vicariance speciation (*v*), 15 sympatric range-copying speciation (*y*) and 9 sympatric-subset speciation (*s*). Under the DEC+J model only two dispersals by range expansion were identified, both to the Atlantic Forest, while 16 jump-dispersal events (*j*) were identified, mainly to the Cerrado (C, four jumps) and Atlantic Forest (D, four jumps) (vertical arrows, Fig. 1.3). Under DEC+J we found: 1 vicariance speciation (*v*), 14 *in-situ* speciation (*y*), and 1 sympatric-subset speciation (*s*) events.

All ecoregions appear overdispersed in the tree; the exceptions are the recently colonized Southern Andes (B) and Pampas grassland (F), which are each restricted to a

single clade. However, the mode of dispersal varies dramatically between DEC and DEC+J models, with the first favoring range expansion with subsequent speciation through vicariance and the second favoring jump-dispersal speciation events.

Diversification rates

The lineage-through-time plot (Fig. 1.4A) indicates a departure from constant diversification rate, which is supported by the rejection of the null hypothesis of constant diversification for the *Hypsiboas pulchellus* group (MCCR gamma statistics; $\gamma = -3.85$; $p < 0.01$). Indeed, most variable-rate diversification models had better fit than the constant-rate models (Table 1.1). The LTT plot shows a substantial acceleration in net diversification around 5 MYA above the null expectation (Fig 1.4A), and this acceleration is quickly followed by a decline towards the present. The model that best explained this pattern was the diversity-dependent logistic (DDL) model (Table 1.1). The BAMM analysis did not identify any rate shift among lineages (Fig S5) and the gradual change in diversification rates was mainly explained by variation in speciation rates starting at a rate of ~0.4 and gradually decreasing as a function of the number of species accumulated (Fig 1.4B). Extinction rates were consistently low (Fig 1.4C) and did not seem to strongly influence the diversification pattern. Estimates were also near zero when considered under all three diversity-dependent models (model likelihood: DDL–E = –81.8, DDL+SE = –82.0, DDL+E = –88.5).

DISCUSSION

Range evolution: dispersal and center of origin

The traditional view of species-rich regions as centers of origin in which a lineage originates and mainly diversifies in a single region (Mora *et al.* 2003) was not supported

by our biogeographic analyses. Under both DEC+J and DEC biogeographic models, the center of origin model (M1) was either rejected (by ~3 lnL units) or strongly rejected (by ~10 lnL units), respectively. Even though, under the DEC+J reconstruction many of the deeper nodes are reconstructed as area E (Araucaria Forest).

The interpretation of the relative importance of dispersal and vicariance differs between the DEC and DEC+J models. For the DEC model, we reconstructed six occurrences of dispersal, of which four were to the Atlantic Forest. The dispersal events reconstructed under DEC are interpreted as dispersals due to range expansion, because DEC does not incorporate jump-dispersal (founder events). Following range expansion, a widespread species may divide into allopatric descendant lineages with restricted ranges. In our reconstruction, four of the six dispersal events were followed by vicariance speciation.

Eight vicariance events were reconstructed; six of these involved lineage divergence into one dry and one humid region (e.g., Cerrado vs. Atlantic Forest). As well, five of the eight involved the Atlantic Forest. However, the vicariance speciation is not necessarily due to the formation of a barrier. For example, the movement of species across the complex heterogeneous landscape in the Neotropics can lead to reduction of gene flow among populations, resulting in speciation with no need to invoke geographic barriers or vicariance events, as has been demonstrated for birds (Smith *et al.* 2014).

Under DEC+J, only two dispersal events through range expansion and one vicariance event (at the root) were identified. Jump-dispersal accounted for 16 of the range-shifts, consistent with constraints of the DEC+J model. These movements, especially between the Cerrado and Atlantic Forest, suggest a series of ecological speciation events through habitat shifts (Rundle & Nosil 2005). Under this model, the ecoregions effectively function as ecological islands of suitable habitat among which a

lineage disperses over time. Thus, our analysis recovers substantial interchange between contrasting habitats since the Miocene.

Species were also generated *in situ*. Several cases of *in situ* speciation in each ecoregion (except for the recently colonized Southern Andes) were reconstructed (under DEC, 15 sympatric range-copying speciation and 9 sympatric-subset speciation; under DEC+J, 14 sympatric range-copying and 1 sympatric-subset speciation events). Overall, speciation occurred through a combination of dispersal and *in-situ* generation, with neither one strongly predominating. Therefore, our study indicates that Neotropical species-rich regions may act simultaneously as *in-situ* generators and accumulators of local species diversity.

We also found strong support for stepping-stone dispersal in the *Hypsiboas pulchellus* group. Stepping-stone is an important dispersal mode among discrete habitats such as islands (Thornton *et al.* 2002); however, we show here that it may apply to large continental settings as well. A stepping-stone dispersal mode was also used to effectively explain the dispersal pattern in poison dart frogs in contiguous regions of South America (Santos *et al.* 2009).

Temporal diversification patterns

A constant diversification rate was rejected for the *Hypsiboas pulchellus* group; instead we found a pattern of diversity-dependence diversification (Moen & Morlon 2014) with speciation rates decreasing as species diversity increases. Although declining, the relatively high rates of speciation (initial value of ~0.4) associated with a constant low extinction rate seem to conform to the expectation of tropical regions as models of both cradle and museum of biodiversity (Jablonski *et al.* 2006; McKenna & Farrell 2006; Moreau & Bell 2013).

Diversity-dependent diversification is a common pattern in large phylogenies (Phillimore & Price 2008; Rabosky & Lovette 2008). However, studies of diversification patterns in the continental Neotropical region have yielded mixed results. While some studies show striking examples of nearly constant diversification (Derryberry *et al.* 2011; Bryson *et al.* 2014; Castroviejo-Fisher *et al.* 2014; Price *et al.* 2014), others show patterns of rate-variable diversification, including diversity-dependent diversification (Weir 2006; Fine *et al.* 2014), and rate shift among lineages (Santos *et al.* 2009; Ramírez *et al.* 2010; Moreau & Bell 2013). Our study accrues more evidence for diversity-dependent diversification pattern in the continental Neotropical region.

One common biological explanation for diversity-dependent diversification is that speciation slows as ecological opportunities and geographic space limits clade growth (Phillimore & Price 2008). To explain the departure from this pattern in the Neotropics, Derryberry *et al.* (2011) have suggested that constant lineage divergence in tropical continental radiations may not be as limited by ecological opportunities as in islands or temperate regions. Although this might be the case, examples of diversity-dependent diversification in the Neotropics suggest that some clades are close to carrying capacity and experiencing constrained ecological opportunities by niche saturation (Weir 2006; Fine *et al.* 2014, and this study). Some other explanations for this pattern include: sampling bias (Cusimano & Renner 2010), environment-driven pulses of speciation (Moen & Morlon 2014), and the effect of geographic mode of speciation (Pigot *et al.* 2010).

Sampling bias is problematic when less than 80% of species are sampled (Cusimano & Renner 2010), which is not the case in the present study; 82% of the species are included. We also expect that the missing species are well dispersed across the phylogeny, based on their distributions and presumed relationships to sampled

species. Moreover, our inclusion of potential cryptic species reduces the bias of undersampling actual diversity. Therefore, we doubt that the slowdown pattern was caused by under-sampling.

The LTT plot (Fig. 1.4A) shows prominent accelerated diversification between 5–10 MYA, which might be explained by an environment-driven pulse of speciation caused by a rapid environmental or geological change (Moen & Morlon 2014), followed by a subsequent slowdown. The increased ecological opportunity provided by the several colonization events to different habitats in South America around this time might have facilitated the increased diversification in this treefrog lineage. Then, subsequent *in-situ* speciation in each region could have promoted rapid saturation of ecological niches, resulting in the diversification slowdown; however, data on niche occupancy and overlap would be necessary to test this hypothesis.

The geographical context of diversification alone might also be responsible for the slowdown pattern in the treefrogs. Pigot *et al.* (2010) demonstrated that allopatric speciation might produce a slowdown in diversification in the absence of any ecological explanation. When speciation occurs mainly via dispersal, as in our DEC+J analysis, the formation of small peripheral isolates leads to a decline in the average range size, reducing the probability of further speciation (Pigot *et al.* 2010). However, we are not able to distinguish between this possibility and niche saturation as an explanation for the slowdown in diversification.

The role of geological climatic changes

The *Hypsiboas pulchellus* group originated in the middle Miocene, a dynamic period marked by major climatic changes. Following the Miocene Climatic Optimum (~17–15 MYA), a period of considerable warming in mid-latitudes, global temperatures

dropped considerably as the Antarctic ice sheet expanded (Zachos 2001). This period of cooling from ~15–13 MYA is known as the Middle Miocene Climate Transition (MMCT) and was characterized by increased aridification and development of savannas and grasslands in mid-latitude regions of South America (Flower & Kennett 1994). It largely overlaps with the initial diversification of this treefrog lineage (Fig. 1.3) during progressively cold and dry conditions, and their establishment in dry open vegetation and cold montane habitats.

The Miocene was also a period of intense geologic activity in South America. The orogeny and rapid uplift of the Andes had major consequences for the climate (Garreaud 2009) and fluvial drainage systems of the entire continent (Potter 1997). The Central Andes experienced the most significant uplift 10–6 MYA (Garzione *et al.* 2008), which preceded the diversification of Clade B (Fig. 1.2) in the Andes starting ~11 MYA. As observed by Köhler *et al.* (2010), the allopatric or parapatric distribution of sister species in the Andes suggests successive splitting events along the Andean latitudinal gradient.

The uplift of the Andes has also had profound effects on climatic patterns of South America, affecting atmospheric circulation and rainfall patterns (Garreaud 2009). It is thought to have contributed to an increase in arid conditions in several parts of the continent, which in association with the climate cooling condition and increased seasonality favored the expansion of grasslands and open vegetation habitats in South America (Flower & Kennett 1994; Garreaud 2009). The expansion and dominance of C_4 grasslands during the Late Miocene and Pliocene had an enormous impact in the expansion of savanna and grassland biomes worldwide (Edwards *et al.* 2010). This may have enabled range expansion of species mostly associated with open/dry habitats, favoring their subsequent speciation. The expansion and dominance of C_4 grasses in South America (Latorre *et al.* 1997; Edwards *et al.* 2010) is predicted to have occurred

from 3–7 MYA, consistent with the colonization of the Pampas grassland and the open-dry vegetation of the Southern Andes and with most speciation events in the Cerrado. The recent radiation of endemic Cerrado angiosperms, for example, has been attributed to the impact of C_4 grassland expansion on the establishment of the modern savanna (Simon *et al.* 2009) and climate cooling during the MMCT (Antonelli *et al.* 2010). Our study further suggests that both events might also explain the diversification dynamics of the fauna associated with open vegetation habitats in South America.

Treefrogs support the recent origin of the Cerrado savanna biota

The Cerrado is a unique savanna in South America with high endemicity level for plants (44%; Klink & Machado 2005), frogs (53%; Valdujo *et al.* 2012) and lizards (39%; Nogueira *et al.* 2011), among others. Contrasting hypotheses concerning the timing of the origin of this savanna have been proposed, ranging from an old Cretaceous origin (Ratter *et al.* 1997) to a more recent origin in the Miocene (~10 MYA) (Simon *et al.* 2009). The hypotheses for an old origin advocate the establishment of highly endemic fauna and flora by long-term *in-situ* diversification. Conversely, hypotheses of recent origin propose diversification of endemic Cerrado vegetation through adaptive habitat shifts from surrounding forest biomes promoted by the recent grassland expansion in South America (Simon *et al.* 2009). Estimates of the time of origin of Cerrado endemics are few, but the existing evidence supports a recent Cerrado origin (Domingos *et al.* 2014; Santos *et al.* 2014). Here, we show that the origin of the endemic Cerrado treefrogs occurred in ~15 MYA, but most diversification events took place in the last 10 MY, supporting recent Cerrado diversification (Simon *et al.* 2009). The endemic Cerrado treefrogs comprise several lineages that are phylogenetically overdispersed, some of which result from recent in-situ diversification (e.g., *H. botumirim* and *H. cipoensis*) and

others from vicariance speciation from forest ecoregions (e.g., Atlantic and Araucaria Forests).

Phylogenetic and taxonomic implications

Our phylogeny generally agrees with previous studies (Faivovich *et al.* 2004; 2005; Köhler *et al.* 2010; Pyron & Wiens 2011). However, our increased taxon sampling reveals two significant departures, highlighting a complex evolutionary and biogeographic history. First, we showed that species of the *Hypsiboas polytaenius* group, defined by the striped dorsal pattern (sensu Cruz & Caramaschi 1998), do not form a single clade (Fig. 1.2C-E). The dorsal pattern in these treefrogs is homoplastic and likely the result of parallel evolution or convergence. Previous phylogenetic studies have used only three closely related species of striped treefrogs (Faivovich *et al.* 2004; 2005). Second, in agreement with Köhler *et al.* (2010), our study does not support the monophyly of Central and Southern Andean species, opposed to the conclusion of previous studies (Duellman *et al.* 1997; Faivovich *et al.* 2004; Faivovich *et al.* 2005; Wiens *et al.* 2010; Pyron & Wiens 2011), some of which were based on largely overlapping datasets.

The topology we recovered has also important taxonomic implications. First, *Hypsiboas phaeopleura* appears nested within the clade formed by populations of *H. goianus* suggesting that *H. phaeopleura* is not a valid species because the name goianus has priority. However, this taxonomic issue is currently under study by the authors.

Additionally, samples allocated to the nominal species *H. polytaenius* belong to three distinct lineages in our tree (Fig. 1.2C). Because its type locality was originally given as “Brazil” by Cope (1869), it is not clear which of the three lineages should bear the species name. Cruzand Caramaschi (1998) narrowed down the type locality to the

region between Rio de Janeiro-RJ and Lagoa Santa-MG, which passes near Petrópolis-RJ. This area encompasses the distribution of two distinct lineages, one with a single individual from Petrópolis-RJ and another containing individuals from eastern Minas Gerais (Supporting Information), making taxonomic decisions based solely on geographic distribution difficult. This taxonomic issue is also under study by the authors.

The third clade in which the name *H. polytaenius* occurs includes individuals with three different names: *H. beckeri*, *H. latistriatus*, and *H. polytaenius*. The average uncorrected genetic distance for 16S among all individuals in this clade is 0.5% (range 0.1–0.7%), a value far smaller than that (3%) recommended for identification of cryptic frog species (Fouquet *et al.* 2007). Therefore, we consider these individuals to belong to the same taxonomic species. Which name is applicable? Given the restriction of the type locality of *H. polytaenius* by Cruzand Caramaschi (1998), the application of the name *H. polytaenius* to this clade is not appropriate because it lies well outside the region of type locality (see above). The names *beckeri* and *latistriatus* were coined in the same paper (Caramaschi & Cruz 2004) and we sampled individuals from both type localities (Poços de Caldas-MG for *H. beckeri*, and Itamonte-MG for *H. latistriatus*). Acting as First Reviser, we choose *Hypsiboas latistriatus* as the valid name for the species (*Hyla beckeri* Caramaschi and Cruz 2004 is a junior subjective synonym of *Hyla latistriata* Caramaschi and Cruz 2004). Given the difficulty in distinguishing the species by the dorsal pattern, documenting their complete distributions will require additional sequence data

Table 1.1: Models of diversification rate in *Hypsiboas pulchellus* group based on AIC ordered by best fit.

#	Model	Model type	Param	lnL	AIC	dAIC
1	**DDL**	**rate-variable**	**2**	**-0.308**	**4.615**	**0.000**
2	yule3rate	rate-variable	5	2.465	5.070	0.455
3	yule4rate	rate-variable	7	4.333	5.333	0.718
4	yule2rate	rate-variable	3	-0.299	6.599	1.984
5	DDX	rate-variable	2	-6.155	16.310	11.695
6	SPVAR	rate-variable	3	-5.551	17.102	12.487
7	BOTHVAR	rate-variable	4	-5.534	19.068	14.453
8	pureBirth	rate-constant	1	-9.537	21.074	16.459
9	birth-death	rate-constant	2	-9.537	23.074	18.459
10	EXVAR	rate-variable	3	-9.569	25.139	20.524

Figure 1.1: Biogeographic models tested in BioGeoBEARS for the evolution of the *Hypsiboas pulchellus* group in South America. Uppercase letters refer to ecoregions: A= Central Andes (5 spp); B = Southern Andes (2 spp); C = Cerrado (6 spp); D = Atlantic Forest (8 spp); E = Araucaria Forest (12 spp); F = Pampas (2 spp). The numbers in the map represent the dispersal steps between ecoregions. The dispersal matrix used in each model is also shown.

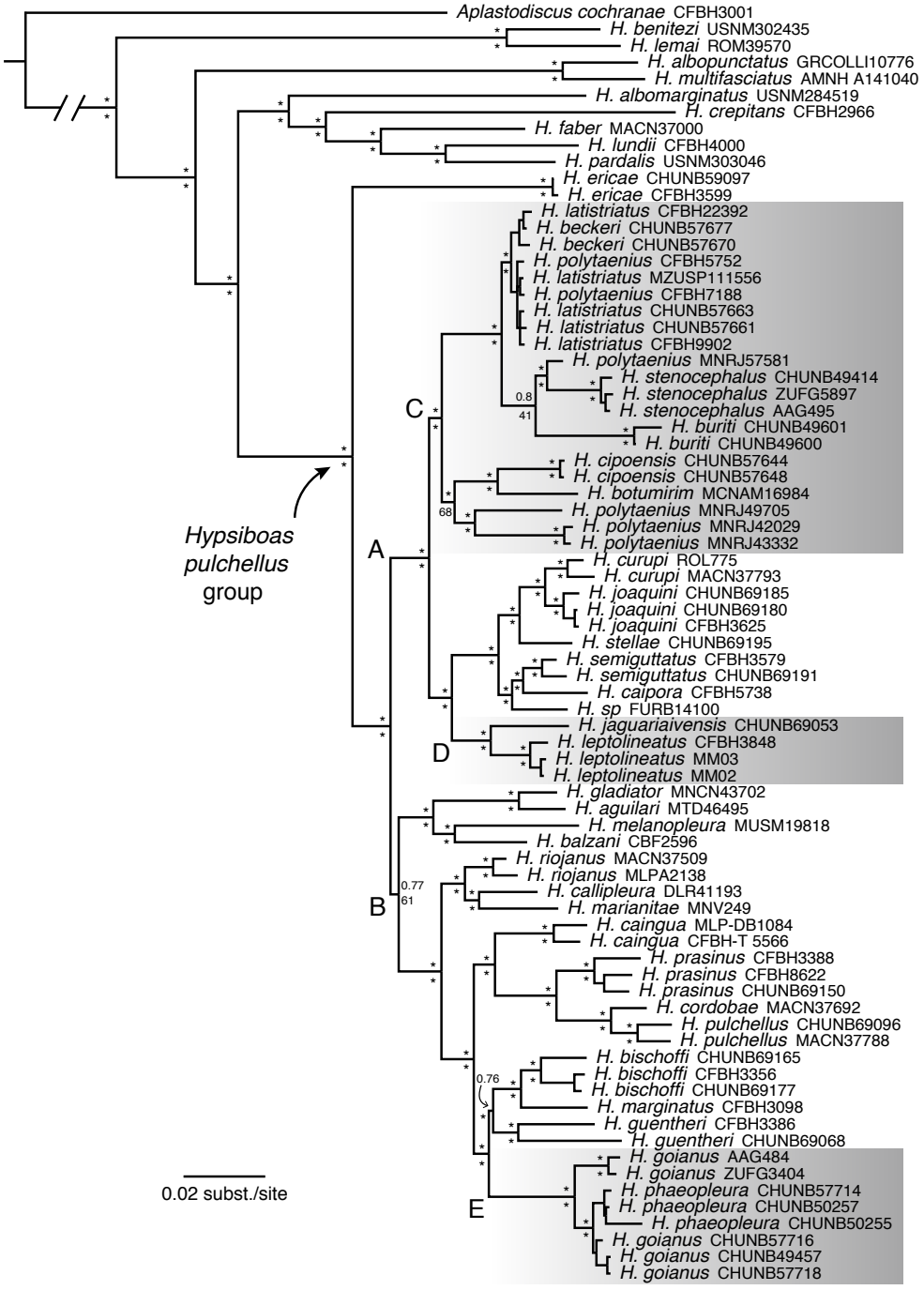

Figure 1.2: Bayesian phylogenetic tree of *Hypsiboas pulchellus* species group. The grey boxes identify the species previously included in the *H. polytaenius* group. Posterior probabilities are given above branches (* indicate values >0.95) and bootstrap support below branches (* indicate values >80). Support for intraspecific relationships are not shown.

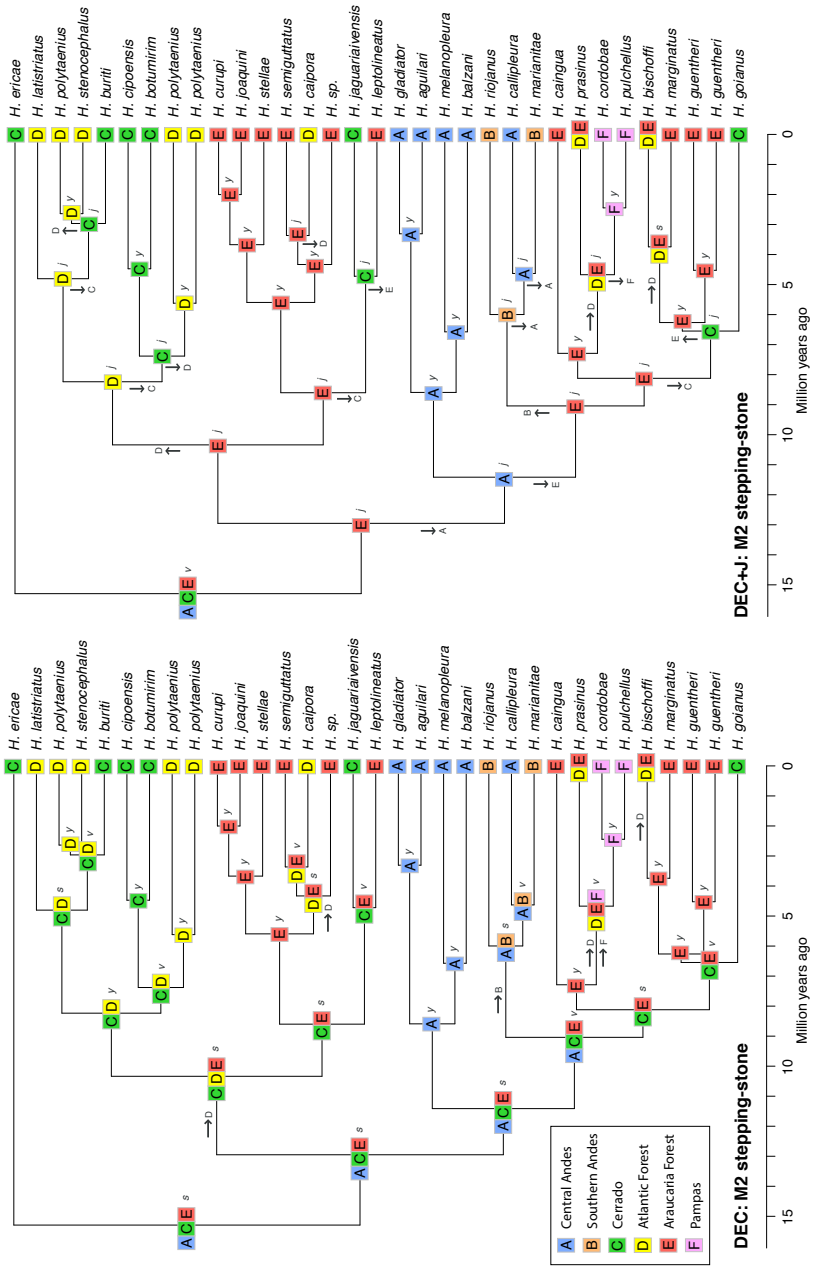

Figure 1.3: Ancestral range reconstruction using the DEC model (left) and the DEC+J model (right) under the favored dispersal model (M2 stepping-stone) for *Hypsiboas pulchellus* group. Dispersal events are displayed as horizontal arrows along branches and jump-dispersal events as vertical arrows at the nodes. Lower case letters refer to speciation modes: *v* = vicariance speciation, *y* = sympatric range-copying speciation, *s* = sympatric-subset speciation, *j* = jump-dispersal speciation only for the DEC+J model.

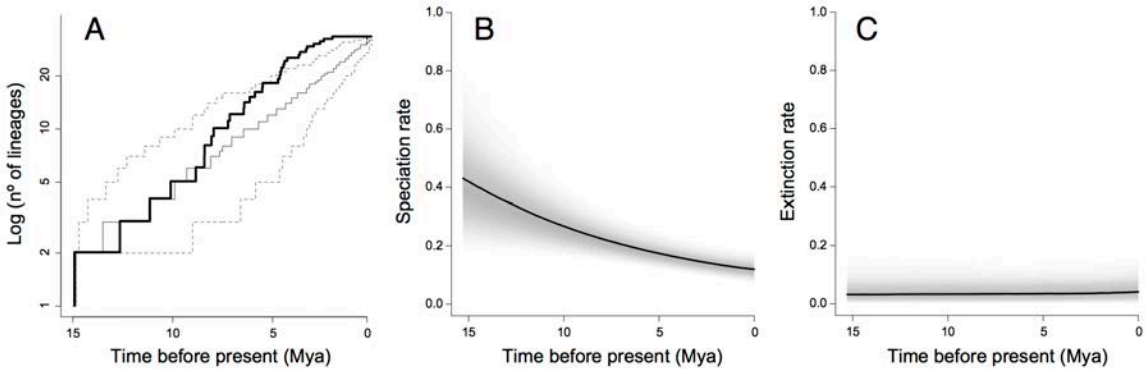

Figure 1.4: (A) Lineage-through-time plot obtained from the chronogram of the *Hypsiboas pulchellus* group showed as a thick line. Expectations from constant diversification obtained through simulations are showed in thinner lines including the 95% confidence interval (dashed lines). (B) Speciation rates estimated over time in BAMM. (C) Extinction rates estimated over time in BAMM. Both B and C include the 95% Bayesian credible regions with higher shading density denoting higher confidence on the estimated rate.

Chapter 2: Species limits and cryptic speciation in Neotropical treefrogs of the Cerrado savanna: integrating evidence from genetics, morphology and mating calls

ABSTRACT

Molecular tools such as coalescent methods of species delimitation have substantially enhanced our ability to determine species boundaries in a more objective way. Groups of organisms with conserved morphology and problematic taxonomy, such as treefrogs, can benefit from these methods, especially in hotspots of biodiversity, where cryptic diversity might be high and underestimated. In this study, we evaluate species limits using an integrative approach combining multi-locus genetic data, morphology, and mating calls for endemic treefrogs of the Cerrado savanna of South America, a hotspot of biodiversity. We test a specific set of hypotheses for the species limits in this lineage with the purpose to solve a taxonomic incongruence recovered by non-reciprocally monophyletic species. We recovered three main geographically delimited clades (*Hypsiboas phaeopleura*, *H. goianus* North, *H. goainus* South), which were consistently delimited as three different species using Bayesian coalescent species delimitation method. Although morphologically divergent, the three recognized lineages showed a great level of morphological overlap, making the classification of individuals based on morphology only a difficult task. A smaller degree of overlap was documented for the acoustic parameters among lineages. The morphological and acoustic differentiation was congruent with the phylogenetic distance among these lineages, highlighting the cryptic status of recent allopatric speciation events, which might require more time to accumulate enough morphological or behavioral diagnostic characters to set the species apart.

INTRODUCTION

In recent years coalescent species delimitation methods have become a powerful tool to investigate phylogeographic patterns and species limits (Fujita *et al*. 2012). These methods can help elucidate diversification patterns in morphologically conservative groups (Satler *et al*. 2013), understand the biogeographic patterns involved in the speciation process (Leaché & Fujita 2010; Leaché *et al*. 2014), and provide insights into the demographic history of the species (Camargo *et al*. 2012). One of the greatest contributions of these methods is the increased objectivity they can provide to taxonomic decisions, since they rely on the coalescent process and not on the researchers subjective impressions about the evolution of the group (Carstens *et al*. 2013; Leaché *et al*. 2014). Hence, the use of coalescent species delimitation combined with other lines of evidence, such as morphology (Pepper *et al*. 2013; Domingos *et al*. 2014; Solís-Lemus *et al*. 2015), is an important step towards an objective integrative taxonomy (Padial *et al*. 2010; Edwards & Knowles 2014).

The recognition of cryptic species in well-studied species groups is not uncommon (Oliver *et al*. 2009; Oliver *et al*. 2010), and this issue is even more prominent in biodiversity hotspots (Scheffers *et al*. 2012). One of the first steps in biodiversity conservation is the proper delimitation and recognition of species, an activity that can be enhanced if evolutionary aspects of the species are known (Diniz-Filho *et al*. 2013). Therefore, the discovery of cryptic lineages can directly contribute to conservation planning (Niemiller *et al*. 2013), particularly so in biodiversity hotspots such as the Cerrado savanna in South America (Silva *et al*. 2014).

Recent phylogeographic studies in the Cerrado have allowed an unprecedented discovery of cryptic lineages in this region (Prado *et al*. 2012; Werneck *et al*. 2012a; Domingos *et al*. 2014; Guarnizo *et al*. 2016), reinforcing the very high levels of

endemism in this savanna (Myers *et al*. 2000). A recent study on the evolution of a radiation of treefrog species comprising several endemics of the Cerrado savanna (Vasconcellos *et al*. *in review*) revealed these can arise from either lineage splitting inside the Cerrado or from habitat shifts with lineages splitting due to movement into savanna habitat from neighboring forested regions (e.g. Atlantic and Araucaria forests). An unexpected relationship between two endemic species of this savanna that is further related to a clade of the Araucaria and Atlantic forest species stimulated the present investigation. The species *Hypsiboas goianus* and *Hypsiboas phaeopleura* appear as non monophyletic (Vasconcellos *et al*. *in review*), with *H*. *phaeopleura* nested within *H*. *goianus*, splitting the latter into two separate geographic clades.

To further investigate species limits, phylogeography and speciation within this lineage, we used an integrative approach, gathering evidence from multi-locus genetic data, morphology, and behavioral traits (i.e., male mating calls). Using a coalescent species delimitation method, phylogenetic reconstructions and series of multivariate statistical analyses, we address the following hypotheses: (1) *Hypsiboas phaeopleura* and *Hypsiboas goianus* are not distinct species [the one-species hypothesis]; (2) *H*. *phaeopleura* is not a valid species, but together with a northern *H*. *goianus* clade, it comprises a distinct species from the southern *H*. *goianus* clade [the two-species hypothesis]; (3) *H*. *phaeopleur*a is a valid species and *H*. *goianus* consists of two distinct cryptic lineages, north and south [the three-species hypothesis]. In addition, we evaluated how morphology and mating calls distinguish different lineages, and whether speciation occurred in the presence of morphological and/or acoustic differentiation.

MATERIALS AND METHODS

Geographic sampling and DNA sequencing

We sampled 34 individuals of *Hypsiboas goianus* distributed in six populations in the core region of the Cerrado savanna and 10 individuals of *Hypsiboas phaeopleura* from the only known locality (the type locality) where this species occurs (Fig 2.1). We also sampled two individuals of *Hypsiboas bischoffi* and *Hypsiboas prasinus* and one individual of *Hypsiboas guentheri* to use as outgroup species in the analyses (Supplementary Table 2.1). Genomic DNA of each sample was extracted from preserved liver or muscle using the DNeasy extraction kit (Qiagen, Inc., Valencia, CA). We sequenced a mitochondrial region of 2424 base pairs (bp) using a set of overlapping primers (Table 2.1) that encompass the mitochondrial genes 12S rRNA, $tRNA^{Val}$, and 16S rRNA (hereafter referred to as 12S-16S), in addition to three nuclear loci: 898 bp of Rhodopsin (2 exon regions + intron), 496 bp of POMC (coding region), and 589 bp of Loc108 (anonymous locus).

Standard PCR reactions were performed with TaKaRa Ex Taq® polymerase kit (Clontech laboratories Inc., CA) using the primers specified in Table 2.1. The following standard thermocycler protocol was used for 12S-16S: initial denaturation at 94ºC for 2 min, followed by 35 cycles (of 30 sec denaturation at 94ºC, 30 sec annealing at 48ºC, 1 min extension at 72ºC), and a final extension phase of 7 min at 72ºC. Conditions for Rhodopsin, POMC, and Loc108 were identical to the above except for the annealing step (Table 2.1). The annealing touchdown protocol was performed for POMC and Loc108 with annealing temperatures starting at 58ºC decreasing 2ºC at each two cycles until reaching 48ºC (for the first ten cycles), then performing another 25 cycles with annealing temperature at 48ºC for 30 sec. Amplified PCR products were purified using Genelute PCR clean-up (Sigma-Aldrich, MO) and sequenced at both directions at Macrogen Inc.

(Seoul, South Korea). We used Geneious® 7.1.7 (Kearse *et al*. 2012) to edit and assemble contiguous sequences for each individual. GenBank accession numbers and voucher information for all sequences are provided in Supporting Information.

Phylogenetic analyses

We performed a multiple alignment for each locus across all individuals using MUSCLE v. 3.8.31 (Edgar 2004) available at (http://www.ebi.ac.uk/Tools/msa/muscle/) EMBL-EBI Web Services. The combined dataset comprising mitochondrial and nuclear sequences consisted of 4407 bp with all loci present in all individuals. First, we performed maximum likelihood phylogenetic analyses on each individual gene using RAxML v.8 (Stamatakis 2014) with the GTRGAMMA substitution model, following the author's recommendations. Then, we estimated a Bayesian phylogenetic tree for the concatenated dataset using MrBayes v3.2.1 (Ronquist *et al*. 2012). Node support was assessed using non-parametric bootstrap values in RAxML (1000 replicates) and Bayesian posterior probabilities in MrBayes.

We employed PartitionFinder v. 1.1.1 (Lanfear *et al*. 2012) to obtain the best-fit partitioning scheme and substitution models for the concatenated dataset using a greedy heuristic algorithm. The best partitioning scheme with linked branch lengths under BIC was: (1) the 12S-16S mitochondrial genes under the GTR+I+G substitution model, (2) Rhodopsin exon (positions 1, 2, 3) under the JC model, (3) Rhodopsin intron under the K80+I model, and (4) POMC (positions 1, 2, 3) and Loc108 under the HKY+I model. We used the MCMC "model jumping" across the model space by setting nst = mixed in MrBayes v3.2.1. Bayesian settings included two independent runs of four Markov chains with random starting trees for 10 million generations sampled at every 1000 generations, and the temperature parameter set of 0.02. We used default priors for all other

parameters, except the rate prior, which was set as "variable", and the branch length prior, for which we decreased the mean of the exponential prior from 0.1 to 0.01, as suggested by Marshall (2010) to avoid the problem of inaccurate long-branch length, particularly common in phylogenies of partitioned dataset with many short branches (Brown *et al*. 2010). Convergence, stationarity, and appropriate sample sizes (ESS > 200) were assessed in Tracer v.1.6 (Rambaut & Drummond 2013). A burn-in of 20% of samples was excluded prior to generating a majority-rule tree.

Lastly, we estimated a species tree in BEAST 2 (Bouckaert *et al*. 2014), using the *BEAST (Heled & Drummond 2010) coalescent method for our multilocus dataset. The topology generated by MrBayes identifies three basal clades, challenging the current taxonomy (see results). Therefore, we used this topology for the *a priori* assignment of individuals into groups, which is necessary for the *BEAST analysis. We used two individuals of *Hypsiboas bischoffi* as the outgroup. The following partitioning scheme and substitution models were selected in PartitionFinder using unlinked branch lengths: (1) 12S-16S under the GTR+I+G, (2) Loc108 under the GTR+I+G, (3) Rhodopsin + POMC under the GTR+I. We ran *BEAST under a relaxed lognormal clock for all partitions and a constant species tree population size for 10^8 generations sampled at every 50,000 generations with 10% burn-in until appropriate ESS values was achieved (>200).

Haplotype network and genetic distances

We used PHASE 2.1.1 (Stephens *et al*. 2001; Stephens & Donnelly 2003) to estimate haplotypes for the three nuclear loci in using the seqPHASE web-server (Flot 2010) to set up input files from fasta alignments. Due to the limitations of the software in identifying haplotypes given length variant heterozygotes, and the limited confidence in genotyping an indel region due to conflicting chromatograms signal, we deleted a region

of 23 base pairs in the alignment of the Rhodopsin consisting of an indel in the intron region. To generate haplotype networks for the phased nuclear haplotypes and mitochondrial haplotypes, we used Haploviewer (Salzburger *et al*. 2011), using as an input the maximum likelihood trees generated in RAxML from the phased nuclear loci and mitochondrial genes.

In addition, we calculated uncorrected sequence divergences (p-distances) among all populations using MEGA (Tamura *et al*. 2011). Then, we evaluated if isolation by distance might explain the pattern of population divergence in the Cerrado savanna. We plotted the pairwise between-population p-distances against the pairwise Euclidean geographic distances; the latter was quantified using the R package *raster* and the geographical coordinates of each population.

Coalescent species delimitation

We used BPP v3.2 (Yang & Rannala 2014) to test species limit hypotheses among the two *H. goianus* groups (north and south) and *H. phaeopleura* (see Results section 3.1). In contrast to previous BPP versions (Yang & Rannala 2010; Rannala & Yang 2013), BPP v3 can simultaneously estimate a species tree under a reversible-jump MCMC species delimitation algorithm, eliminating the concern about over-estimating the number of species (Leaché & Fujita 2010; Caviedes-Solis *et al*. 2015). BPP v3 estimates a species tree using a Subtree Pruning and Regrafting (SPR) algorithm (Rannala & Yang 2015) under the multi-species coalescent model, concurrently testing species hypotheses by collapsing the branches of the different possible phylogenetic hypotheses (species-trees) and comparing their posterior probabilities.

The three putative species were: *H. goianus* North, *H. goianus* South, and *H. phaeopleura* (Fig. 2.2). Since BPP tests all possible species-tree hypotheses (i.e., all

possible topologies), it also tests the hypotheses that any combination of two species might actually form one species only, and that all three species might form a single species. We tested different gamma priors for population size parameters (θs) and for the age of the root in the species tree (τ_0), to represent different possible speciation histories: (1) large population size and deep divergence – we used G (1, 1000) for θs and G (2, 1000) for τ_0; (2) small population size and shallow divergence – G (2, 100) for both priors; and (3) small population size and deep divergence – G (1, 100) for θs and G (1, 1000) for τ_0. The Dirichlet prior (Yang & Rannala 2010) was used for other divergence time parameters.

We ran analyses for 5 x 10^5 MCMC generations, with sampling frequency of five and using 10^4 burn-in generations. We used both available reversible-jump MCMC species delimitation (Yang & Rannala 2010) algorithms, either excluding or leaving the gaps in the alignment (cleandata = 1, the program remove all columns that have gaps or ambiguity characters, and cleandata = 0 means that these will be used in the likelihood calculation). To evaluate convergence for each analysis type we conducted at least two independent runs starting with random tree topologies.

Morphometric analyses

We also performed morphometric analyses of the preserved treefrogs to evaluate whether the lineages identified by the above molecular analyses are morphologically differentiated, and whether divergences among individuals and populations are congruent with the molecular patterns. We measured preserved individuals from 6 of the 7 localities sampled in our molecular phylogenetic analyses (except for Silvânia - GO) in addition to 7 other localities from which tissue samples were not obtained, but which are located within the area occupied by one of the three identified lineages. In total, we measured 164

individuals from 13 localities deposited at the Coleção Herpetológica da Universidade de Brasília (CHUNB).

A single person (MMV), using a digital caliper (0.01 mm accuracy), measured all treefrogs with no knowledge on their original population or lineage to avoid any bias. Individuals had their sex determined by visually inspection for the presence of vocal sac and prepollical spine, which are only present in males. We used only males in further analyses to reduce any variation due to the small number of preserved females among populations. Sexual dimorphism in size and shape is widespread among treefrogs (Shine 1979; Woolbright 1983; Guimarães *et al*. 2011), and given that very few females were available to account for sexual variation, we excluded females from morphological analyses. We selected the following 16 morphometric variables to maximize variation among individuals: snout-vent length, head length, head width, snout-eye distance, naris distance, inter-ocular distance, eye diameter, tympanum diameter, forelimb length, forearm length, third finger length, thigh length, shank length, foot length, and fourth toe length, following (Duellman 2001). We selected the right side of the body for measurements.

We first checked and removed eventual outliers in the $\log_{10}$ transformed morphometric data based on the distribution of z-scores for univariate and squared Mahalanobis distance for multivariate outliers (significance level of 0.001), using the mvoutlier R package (Filzmoser *et al*. 2012). To adjust for body size, we used the Burnaby's back-projection method (Burnaby 1966) recommended by Rohlfand Bookstein (1987) and McCoy *et al*. (2006) as an unbiased method that effectively removes size variation correctly quantifying shape variation. However, we did not use the common PC1 as a body size metric as suggested by McCoy *et al*. (2006) and heavily criticized by Berner (2011). To partition the total morphometric variation into size and shape, we first

defined an isometric size eigenvector with values equal to $p^{-0.5}$, where p is the number of variables, in a procedure described by Somers (1986) applying the allometry equation (Jolicoeur 1963). Next, we obtained scores for this eigenvector, hereafter called body size, by postmultiplying the n x p log-transformed character matrix, where n is the number of specimens, to the p x 1 eigenvector. Then, we removed the effect of body size from the log-transformed variables using Burnaby's back-projection method (Burnaby 1966), postmultiplying the n x p log-transformed character matrix to a symmetric p x p matrix (L), defined as:

$$L = I_p - V(V^T V)^{-1} V^T$$

where I_p is a p x p identity matrix, V is the isometric size eigenvector defined above, and V^T is the transposed V matrix (Rohlf & Bookstein 1987). We refer to the resulting size-adjusted variables hereafter as shape variables.

We then conducted analyses of variance to test for differences in body size (ANOVA) and shape (MANOVA) among lineages. We also performed discriminant function analyses (DFA) using the MASS R package (Venables & Ripley 2002) to assess the degree of shape variation among lineages. In addition, an automated AIC model selection in the MuMIn R package was used to evaluate the importance of variables in lineage discrimination. All calculations and statistical analyses were conducted in R version 3.2.1 (R Core Team 2015).

Call analyses

We obtained male advertisement calls from all seven populations sampled for DNA. Calls from 1 to 4 males per population were recorded using a Marantz PMD660 digital recorder and a Senheiser ME67 directional microphone, except for the call from Silvânia-GO, which was obtained from a multimedia archive (Toledo *et al*. 2007). We

analyzed all calls using Raven Pro 1.5 software, available at www.birds.cornell.edu/raven, at a sampling rate of 44 kHz. Two call types were identified, as defined by Menin *et al*. (2004): a trilled call composed of three to six notes and a call composed of two harsh notes, hereafter referred to as call A and call B, respectively. Although, most recorded individuals emitted both calls intermittently, call A was the only one recorded in all individuals. Therefore, we analyzed call A only, measuring two to nine calls (depending on availability) and all notes within each call for each recorded individual to obtain an average individual estimate. The following six measurements were taken: number of notes per call, call duration time (s), average note duration time in a call (s), interval between notes (s), dominant frequency (Hz), and fundamental frequency (90% band width) (Hz). Acoustic variables differentiation among lineages was also tested with a MANOVA and examined using DFA in R (R Core Team 2015).

RESULTS

Phylogenetic analyses

The individual gene trees for 12S-16S, Rhodopsin, POMC, and Loc108 varied slightly from each other, reflecting their different evolutionary rates and amount of lineage sorting among populations (Supp. Fig 2.1). The marker that most clearly sorted lineages into well-defined and strongly supported clades was the 12S-16S. The concatenated MrBayes phylogeny (Fig. 2.2) depicted a relationship in which *Hypsiboas phaeopleura* and *Hypsiboas goianus* are not recovered as reciprocally monophyletic lineages. Although all individuals of *H. phaeopleura* from the single known locality form a well-supported clade, this clade is nested within *Hypsiboas goianus* clade. Our

concatenated topology recovers with very high support (posterior probability = 1) two basal geographically delimited lineages: a southern lineage comprising populations of *H. goianus* from Silvânia-GO, Catalão-GO and Perdizes-MG, and a northern lineage including *H. phaeopleura* and populations of *H. goianus* from Santo Antônio do Descoberto-GO, Brasília-DF and São João d'Aliança-GO. However, to recognize *H. phaeopleura* as a distinct species from *H. goianus*, we separated the northern lineage further into its two distinct clades: the highly supported *H. phaeopleura* (posterior probability = 1) and the weakly supported *H. goianus* North (posterior probability = 0.6). Therefore, for further hypothesis testing, three lineages were defined: *H. phaeopleura*, *H. goianus* North, and *H. goianus* South (Fig. 2.2).

The *BEAST multi-locus coalescent method, confirmed the relationship found in the concatenated analyses (Figs. 2.2 and 2.3). *Hypsiboas phaeopleura* was recovered as the sister lineage of *H. goianus* North with very high support (posterior probability = 0.99). And this clade was then related to the *H. goianus* South lineage, also with high support (posterior probability = 1), making *H. goianus* a paraphyletic species. The level of congruence among all trees, as visualized with DensiTree (Bouckaert 2010) was very high.

Haplotype networks and genetic distances

The mitochondrial genes 12S and 16S are comprised of unique haplotypes for each of the three lineages: *H. phaeopleura*, *H. goianus* North, and *H. goianus* South (Figs. 2.4A-B). The haplotype networks for these two markers support a clear divergence among these lineages, which do not share any haplotypes. The networks showed several mutational steps (i.e. missing haplotypes) between the *H. goianus* South and the haplotypes of the other two lineages, and also few mutational steps separating *H.*

phaeopleura from *H. goianus* North haplotypes (Fig. 2.4A-B). In contrast, the three nuclear markers showed very little genetic structure and high levels of incomplete lineage sorting (ILS) (Figs. 2.4C-E). There are many shared common haplotypes among lineages, and no evidence for a clear geographic pattern of lineage sorting.

Coalescent species delimitation

All BPP runs consistently returned identical results, corroborating our three-species hypothesis, in which each of the three lineages is considered a different species. The three species were always recovered with a posterior probability of 1, and the best coalescent species tree hypothesis had the same topology as the concatenated and species tree phylogenies (Figs. 2.2 and 2.3), i.e. (*H. goianus* South, (*H. goianus* North, *H. phaeopleura*)).

Morphometric differentiation

After removing uni- and multivariate outliers, our total dataset consisted of 153 measured individuals distributed in the three lineages as follows: 77 *Hypsiboas goianus* North, 56 *H. phaeopleura*, and 20 *H. goianus* South (Supplementary Table 2.2). We found a significant difference in body size among the three lineages (ANOVA $F_{2,150}$ = 5.47; p = 0.005), which was mainly due to the significantly larger size of *H. goianus* South compared to the other two lineages (Fig. 2.5). We found a significant difference in shape variation among lineages (MANOVA $F_{2,150}$ = 2.90; p < 0.001). However, the discriminant function analysis (DFA) plot showed a great level of morphological overlap among all three lineages, making the lineage differentiation based solely on morphology a very challenging task (Fig. 2.6). A cross-validation method for classifying individuals based on their DFA scores (100 bootstraps) showed a high misclassification error of 0.34 ± 0.01. When *H. phaeopleura* + *H. goianus* North were treated as a single unit, different

from *H. goianus* South, the misclassification error reduced to 0.15 ± 0.003. Using model selection, eight variables showed an importance level > 0.80 for discriminating lineages: snout-eye distance, naris distance, fourth toe length, inter-ocular distance, eye diameter, head length, forelimb length, and head width (Supp. Fig 2.2).

Acoustic differentiation

No outliers were found in the acoustic data. We measured a total of 64 calls in 13 individuals distributed in the three lineages as follows: five *H. goianus* North, three *H. phaeopleura*, and five *H. goianus* South (Supplementary Table 2.3). The acoustic difference among lineages approached significance (MANOVA $F_{2,10} = 2.54$; $p = 0.057$); the small sample sizes might have precluded discrimination among them. The best call predictors of lineage differentiation were maximum frequency (ANOVA $F_{2,10} = 7.44$; $p = 0.01$) and call duration (ANOVA $F_{2,10} = 3.37$; $p = 0.07$). For both variables, *H. goianus* South showed greater divergence with higher maximum call frequency and shorter call duration. The DFA plot showed a higher level of differentiation in acoustic parameters, with a small overlap among lineages, mostly between *H. phaeopleura* and *H. goianus* North, with *H. goianus* South more acoustically differentiated (Fig. 2.7). A cross-validation (100 bootstraps) showed a misclassification error of 0.53 ± 0.02 for the three-lineage configuration, in which, similar to the morphometric results, was substantially reduced to 0.29 ± 0.02 when considering the two-lineage configuration. Model selection of variable importance could not be performed due to the small sample sizes.

DISCUSSION

Our concatenated and gene tree analyses depicted a relationship in which populations of *H. goianus* are paraphyletic with respect to *H. phaeopleura*. Three primary

lineages were identified: *H. goianus* South, *H. phaeopleura*, and *H. goianus* North (Fig. 2.2). The coalescent species delimitation analysis supported that these lineages comprise three different species. *Hypsiboas phaeopleura* (Caramaschi & Cruz 2000) was originally described from a single locality, Alto Paraíso de Goiás-GO , which is only 53 km apart from *H. goianus* type locality São João d'Aliança-GO (Lutz 1968) (Fig. 2.1). Despite the geographical proximity, there is enough evidence of geographic structure and limited gene flow between these species, especially in the mitochondrial genes (Fig. 2.4). And, in spite of being morphologically cryptic, they can be somewhat sorted based on acoustic components (Fig. 2.7). The recognition of a third cryptic species (*H. goianus* South) was more strongly supported by other lines of evidence, including complete sorting of mitochondrial haplotypes, larger body size and more divergent acoustic parameters such as higher maximum frequency and shorter call duration. The larger acoustic and morphological differentiation distinguishing *H. goianus* South from the other lineages are in agreement with the larger genetic distance of this lineage from the other two.

Call parameters such as maximum frequency can be correlated to body size, but it usually shows an inverse relationship, with larger body size males emitting lower frequency calls (Gingras *et al*. 2012). This negative relationship is based on allometric constraints of the sound production mechanism; in which larger larynx with more robust vocal cords produce lower frequency sounds (Guerra *et al*. 2014). Therefore, body size can be a very good predictor of larynx size, especially with dramatic body size variation, which negatively correlates with call frequency as demonstrated in some comparative studies (Hoskin *et al*. 2009; Gingras *et al*. 2012). *Hypsiboas goianus* South showed larger body size and higher call maximum frequency when compared with the other two lineages, which is in opposite direction with the acoustics constraint. However, the maximum body size difference among those lineages is less than 2mm (Table 2.3), which

might not be large enough to pose a significant physical constraint in the call maximum frequency. Therefore, the divergence of call parameters might be reflecting more the acoustic properties of the calling habitat for each lineage instead of allometric variation.

We found stronger genetic structure in the mitochondrial markers separating the three lineages. In contrast, nuclear loci showed very low levels of genetic differentiation and lineage sorting. This discrepancy regarding the genetic structure among genetic markers might indicate slow rates of evolution in the nuclear loci used in this study compared to the 12S-16S mitochondrial evolutionary rates, which has been estimated as 0.0025-0.0028 substitutions per site per million years for other frog species (Evans *et al*. 2004; Lemmon *et al*. 2007). The low evolutionary rates of the nuclear loci associated with a recent time of lineage split might not have been sufficient to exhibit enough lineage sorting. However, other studies of Hylinae treefrogs using some of the same nuclear genes used in this study found a better resolution in the nuclear loci sorting among recent lineages (Funk *et al*. 2012; Salerno *et al*. 2015).

An alternative possibility is that recent or ancestral gene flow among lineages could be male-biased in these species, which is a common explanation for mito-nuclear biogeographic discordance for amphibians (Toews & Brelsford 2012). Therefore, nuclear markers would appear as almost panmictic, while the maternally inherited mitochondria would indicate greater levels of geographic structure. Very limited amount of male gene flow could result in this discrepancy, especially considering the shorter coalescence time for mitochondrial markers (Zink & Barraclough 2008). However, there is no information about male-biased dispersal in these treefrogs.

The multi-species coalescent model implemented in BPP accommodates incomplete lineage sorting (ILS) (Yang & Rannala 2010), and its species delimitation algorithms can provide correct species delimitations in the presence of migration, as

supported by simulations (Zhang *et al*. 2011) and empirical tests (Camargo *et al*. 2012). Hence, our highly supported coalescent three-species hypothesis seems to be robust, even when considering a speciation scenario with ILS and migration. Although there is a weak degree of differentiation in the morphological data and male calling traits (Fig. 2.4), under the General Lineage Concept (de Queiroz 1998), a complete discrimination is not absolutely necessary for the recognition of the different species.

Despite using several traits including multi-locus sequence data, morphology and mating calls, genetic markers and the use of coalescent methods usually show a greater level of resolution over any other character or method used for species delimitation. Genetic markers typically identify recently divergent lineages more clearly than morphological or behavioral traits (Fujita *et al*. 2012; Sistrom *et al*. 2013), especially when selection for divergence in these traits is weak (as for allopatric or parapatric speciation) and traits are not directly involved in the speciation process (but see Boul *et al*. 2007; Guerra & Ron 2008). When speciation is a recent and ongoing process, insufficient time might have passed to accumulate enough neutral morphological or behavioral differences to completely discriminate among those lineages (de Queiroz 2007).

The discovery of cryptic species in a lineage with a narrow distribution in the central region of the Cerrado raises concern over the conservation status of these species. These stream-breeding treefrogs occur in gallery forests along streams in the Cerrado savanna (Menin *et al*. 2004). But, they only occur in headwater environments of high-elevation plateaus, which are particularly targets for the expansion of agriculture (Silva *et al*. 2006). The accelerated rate of destruction of their natural habitats, their high specialization to stream habitats, narrow distributions, and small clutch sizes (of ~180

eggs for Hypsiboas goianus; Menin *et al*. 2004) make them highly vulnerable to extinction using the metrics of Ribeiro *et al*. (2016).

The first step towards a comprehensive conservation assessment of any species is an accurate taxonomic classification that reflects their evolutionary history (Moritz 1994; Hey *et al.* 2003; Mace 2004) including information about their phylogenetic relationships (Diniz-Filho *et al.* 2013). Moreover, an effective conservation plan requires knowledge about the identity and distribution of the organisms (Margules & Pressey 2000). Coalescent species delimitation methods therefore have the potential to play an essential role in facing the current biodiversity crisis (Fujita *et al.* 2012). These methods will likely be an important tool to reveal the accurate species diversity in important hotspots of biodiversity such as the Cerrado of South America.

Table 2.1: Primers used in this study.

Primer	Direction	Primer sequence (5' – 3')	Source	PCR protocol	Annealing Conditions
12S-16S (mitochondrial)					
MVZ59	forward	ATAGCACTGAAAAYGCTDAGATG	Goebel, 1999	standard	48ºC for 30 sec (x35)
tRNAval_goi	reverse	GGTGTAAGCGAGAYGCTTTRTTTAAG	This study	standard	48ºC for 30 sec (x35)
12s1_goi	forward	CCTAGAGGAGCCTGTCCTATAATCG	This study	standard	48ºC for 30 sec (x35)
16sh_goi	reverse	GTTAAACCATGATGCAAAAGGTAC	This study	standard	48ºC for 30 sec (x35)
12sm_goi	forward	CAGTAGAAGAGGTAAGTCGTAACATG	This study	standard	48ºC for 30 sec (x35)
16_sa	reverse	ATGTTTTTGGTAAACAGGCG	Darst and Cannatella 2004	standard	48ºC for 30 sec (x35)
16sc	forward	GTRGGCCTAAAAGCAGCCAC	Darst and Cannatella 2004	standard	48ºC for 30 sec (x35)
16sd	reverse	CTCCGGTCTGAACTCAGATCACGTAG	Darst and Cannatella 2004	standard	48ºC for 30 sec (x35)
Rhododpsin (nuclear coding region with intron)					
Rhod_1L	forward	AACGGAACAGAAGGCCCAAACTT	Salerno et al. 2015	standard	55ºC for 45 sec (x35)
Rhod_1U	reverse	GCCAAAGCCATGATCCAGGTGA	Salerno et al. 2015	standard	55ºC for 45 sec (x35)
POMC (nuclear coding region)					
POMCF1	forward	ATATGTCATGASCCAYTTYCGCTGGAA	Vieites et al 2007	touchdown	58-50ºC (x10) + 48ºC (x25)
POMCR1	reverse	GGCRTTYTTGAAWAGAGTCATTAGWGG	Vieites et al 2007	touchdown	58-50ºC (x10) + 48ºC (x25)
Loc108 (nuclear anonymous loci)					
Loc108_F	forward	TCATCGCGTCCAAACTTAAT	Lemmon and Lemmon 2012	touchdown	58-50ºC (x10) + 48ºC (x25)
Loc108_R	reverse	CCTTTTCTGGATCATACTCTTCTT	Lemmon and Lemmon 2012	touchdown	58-50ºC (x10) + 48ºC (x25)

Table 2.2: Summary statistics of 16 morphometric variables including mean, standard deviation (sd), minimum (min) and maximum (max) values.

	group: goiS (N = 20)				group: goiN (N = 77)				group: phaeo (N = 56)			
	mean	sd	min	max	mean	sd	min	max	mean	sd	min	max
SVL	34.39	1.12	32.52	36.87	33.82	1.83	30.15	37.46	32.62	1.71	28.18	35.71
HeadL	9.21	0.33	8.82	9.82	9.25	0.51	8.14	10.63	9.13	0.61	7.96	11.18
HeadW	10.07	0.43	9.32	10.69	9.88	0.6	8.7	11.28	10	0.55	8.8	10.93
SnoutL	4.45	0.21	3.9	4.77	4.41	0.33	3.78	5.51	4.34	0.26	3.82	5
SnoutEyeD	2.65	0.14	2.38	2.9	2.52	0.27	1.89	3.44	2.56	0.24	2.05	3.34
NarisD	2.67	0.24	2.23	3.11	2.64	0.23	2.14	3.22	2.78	0.25	2.23	3.4
IOrbD	6.1	0.27	5.61	6.45	5.92	0.44	4.92	6.78	6.02	0.34	5.05	6.45
EyeDiam	3.46	0.17	3.14	3.78	3.51	0.24	3.02	4.12	3.61	0.2	3.19	4.1
TimpD	1.62	0.19	1.31	2.05	1.58	0.14	1.24	1.95	1.53	0.17	1.24	1.84
ForelimbL	6.72	0.5	5.99	7.78	6.23	0.68	4.38	8.1	6.02	0.62	4.59	7.22
ForearmL	16.56	0.57	15.43	17.55	15.87	0.96	13.36	17.95	15.38	0.77	13.65	16.65
Finger3L	7.84	0.63	6.36	8.69	7.38	0.6	6.19	8.67	6.97	0.6	5.8	8.25
ThighL	17.79	1.03	15.22	19.7	17.25	1.25	14.05	20.06	16.53	0.88	14.45	18.34
ShankL	18.47	0.87	17.13	20.23	17.84	1.09	15.72	19.96	17.34	0.96	15.35	19.33
FootL	26.45	0.99	24.75	28.11	25.51	1.64	22.05	28.83	24.86	1.36	21.9	27.25
Toe4L	11.53	0.82	9.66	12.88	10.95	0.93	9.06	13.44	10.74	0.81	9.05	12.31

Table 2.3: Summary statistics of six call variables including mean, standard deviation (sd), minimum (min) and maximum (max) values.

	group: goiS (N = 5)				group: goiN (N = 5)				group: phaeo (N = 3)			
	mean	sd	min	max	mean	sd	min	max	mean	sd	min	max
N° notes	4.28	0.53	3.6	5	4.16	0.74	3.3	5	3.67	0.32	3.3	3.9
Delta time	0.03	0.01	0.02	0.05	0.04	0.02	0.03	0.07	0.03	0.01	0.02	0.04
Call duration	0.29	0.04	0.24	0.33	0.42	0.09	0.32	0.56	0.33	0.12	0.2	0.44
Band width 90	1027.8	469.9	580.0	1757.1	1503.1	1251.0	344.5	3572.5	602.1	272.1	418.0	914.7
Max frequency	3247.7	176.2	2949.2	3410.8	2819.7	163.4	2559.4	2928.5	3048.3	196.5	2908.8	3273.0
Inter-notes time	0.05	0.02	0.03	0.08	0.09	0.04	0.04	0.13	0.09	0.03	0.06	0.12

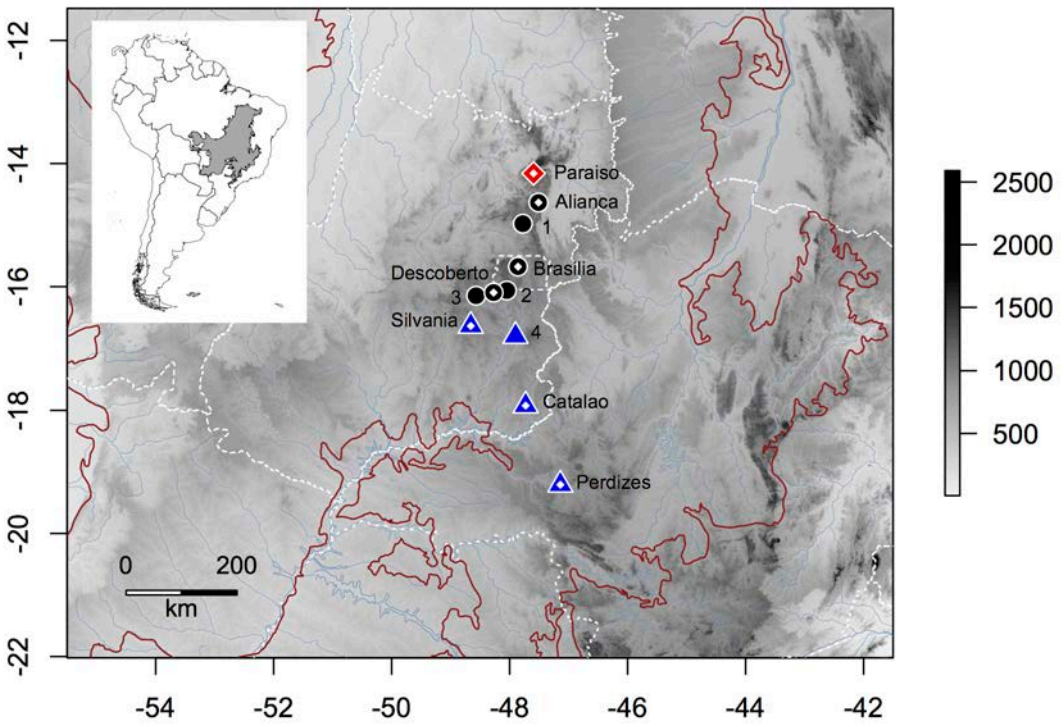

Figure 2.1: Map showing the central region of the Brazilian Cerrado delimited in red, with the locations of sampled populations identified by different symbols that correspond to different lineages: red diamond = *Hypsiboas phaeopleura*, black circle = *H. goianus* North, blue triangle = *H. goianus* South. Symbols with a central white dot represent localities sampled for genetic analyses, which are also followed by names. Numbers correspond to localities sampled in morphological analyses only: 1 = Água Fria de Goiás – GO, 2 = Novo Gama – GO, 3 = Alexânia – GO, 4 = Luziânia – GO.

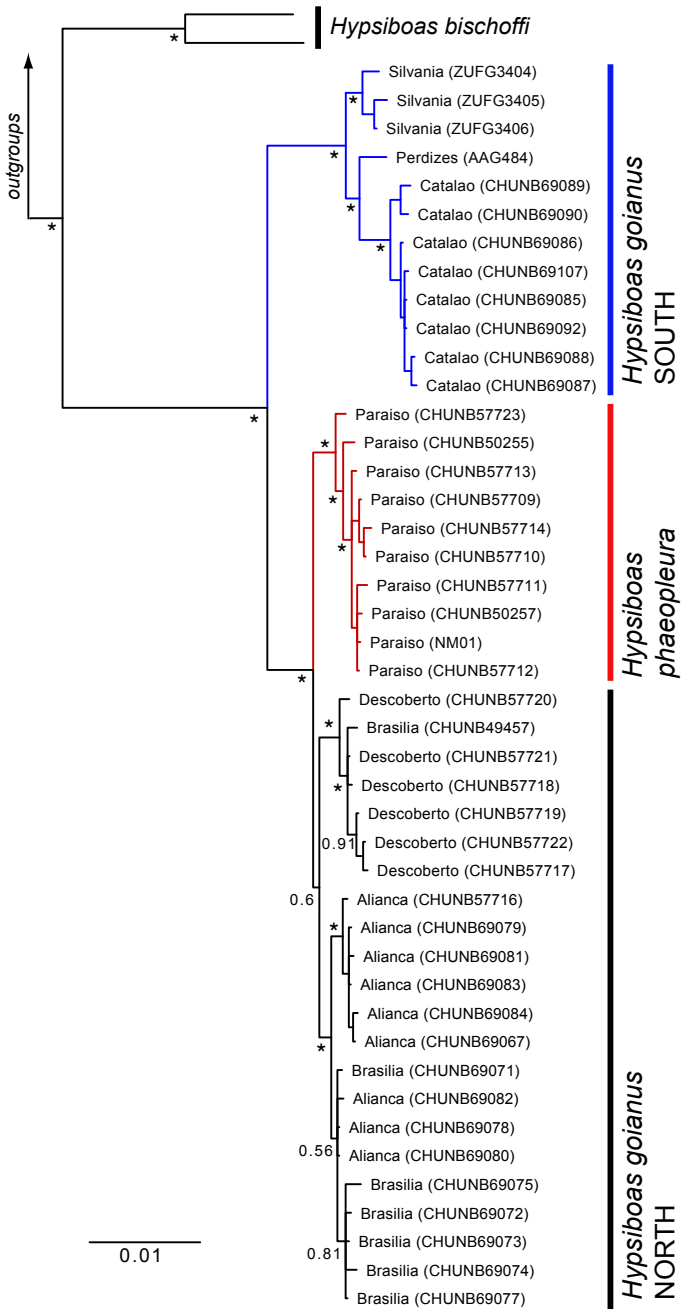

Figure 2.2: Phylogenetic tree obtained in MrBayes using the concatenated dataset of five loci. Bayesian posterior probabilities are shown at the nodes, * represents values > 0.95.

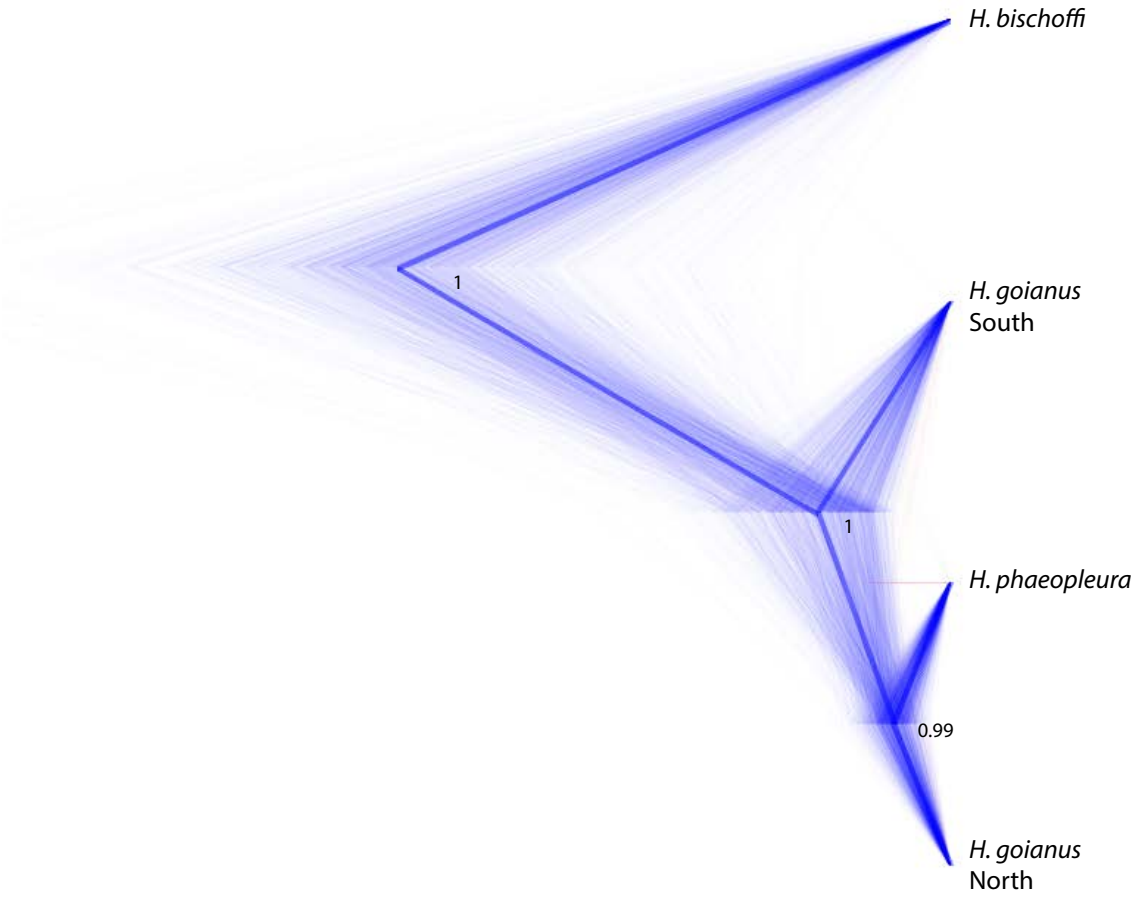

Figure 2.3: Coalescent species-tree generated in *BEAST. DensiTree visualization of all posterior trees with the consensus tree depicted as a thicker line. Posterior probabilities are shown at the nodes.

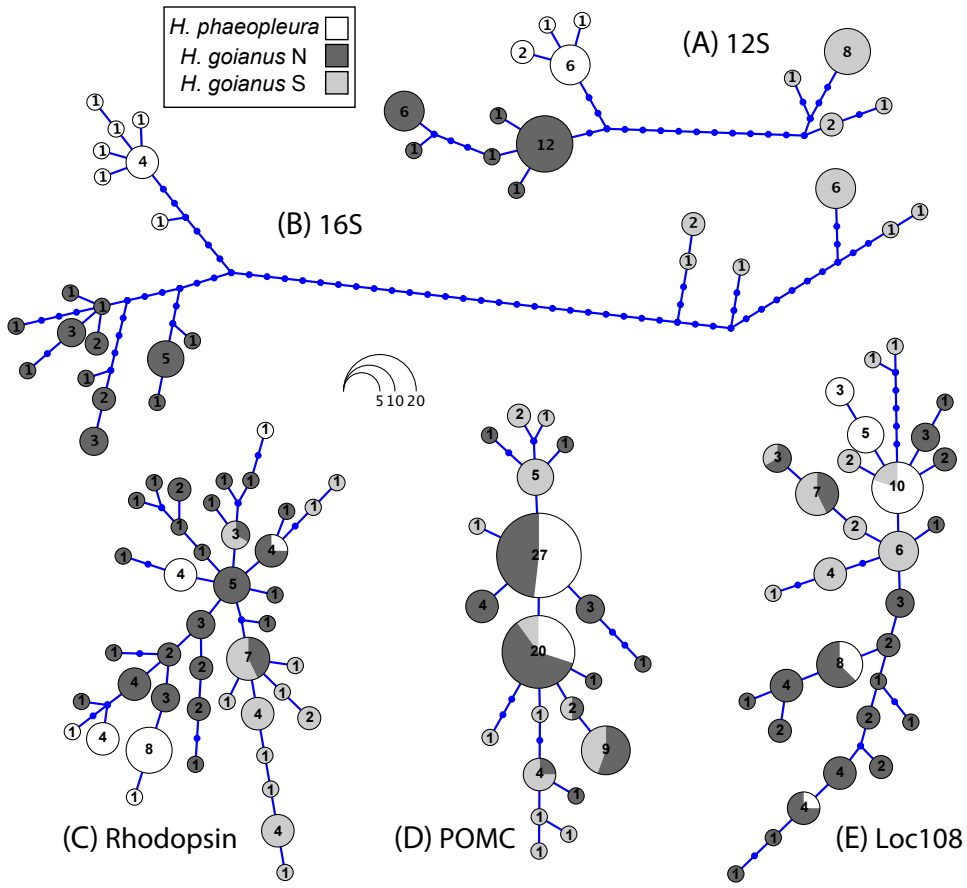

Figure 2.4: Haplotype networks generated in HAPLOVIEWER for the mitochondrial genes (A) 12S, (B) 16S, and for the phased nuclear haplotypes of (C) Rhodopsin, (D) POMC, (E) Loc108. Circles represent different haplotypes scaled to sample size, which is represented by the number inside the circles. Colors correspond to each of the three lineages according to the legend, and dots represent mutational steps between haplotypes (i.e. missing haplotypes).

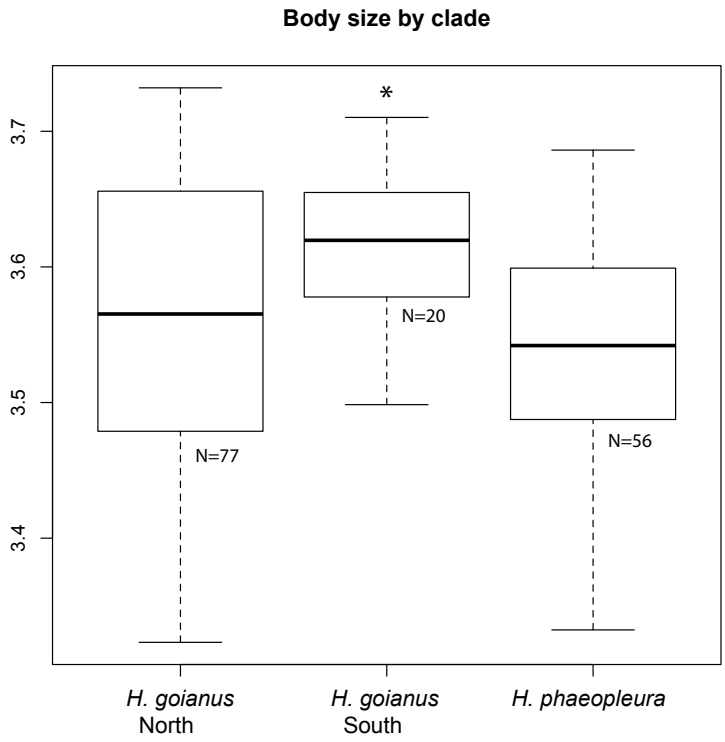

Figure 2.5: Boxplot showing body size variation among the lineages: *Hypsiboas goianus* North, *H. goianus* South, *H. phaeopleura*. Darker line in the middle depicts the median value and boxes delimit the 25% and 75% quartiles. Asterisk represents significant difference ($p < 0.05$) when compared with other lineages.

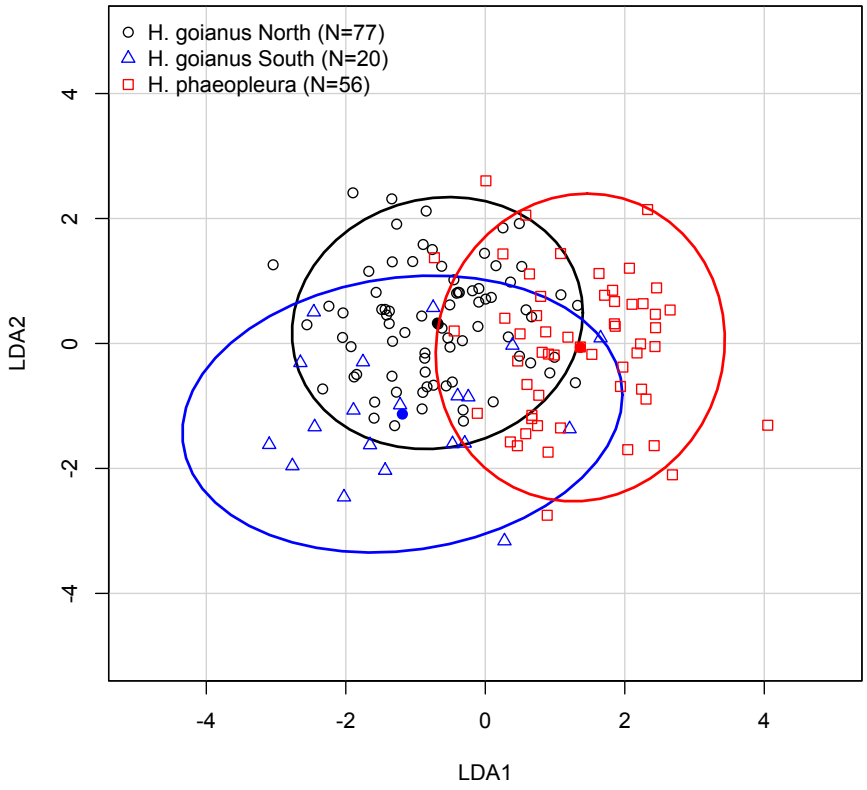

Figure 2.6: Plot of discriminant function analysis (DFA) of the three identified lineages for morphometric shape variables (size corrected) with a 95% confidence ellipse. The symbols (square, circle and triangle) correspond to individuals in each lineage as shown in the legend. The multivariate centroid for each lineage is shown as a filled circle in the middle of each ellipse.

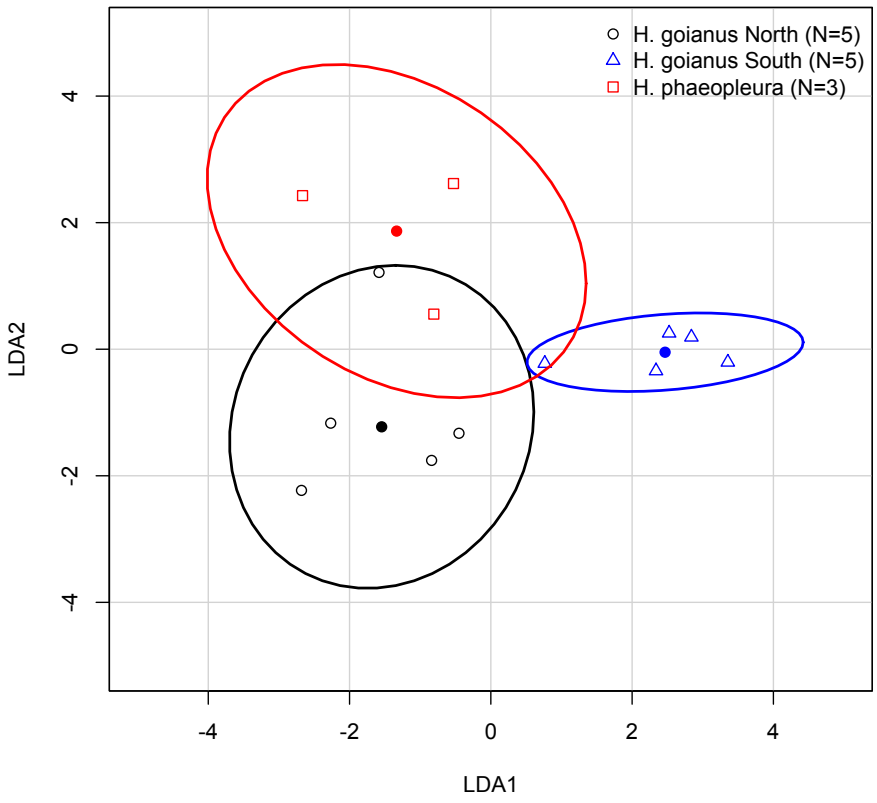

Figure 2.7: Plot of discriminant function analysis (DFA) of the three identified lineages for male advertisement call parameters with a 90% confidence ellipse. The symbols (square, circle and triangle) correspond to individuals in each lineage as shown in the legend. The multivariate centroid for each lineage is shown as a filled circle in the middle of each ellipse.

Chapter 3: Historical climate change shapes population structure and genomic divergence of treefrogs in the Neotropical Cerrado savanna

ABSTRACT

Although the impact of Pleistocene glacial cycles was once dismissed in tropical regions, increasing evidence suggests that tropical organisms have been greatly affected by Pleistocene climatic fluctuations resulting in distributional shifts. However, the genetic consequences of such responses to past climate change are just now being uncovered in several regions. Using genome-wide SNPs and mitochondrial DNA combined with predictions of species distribution models across the late Quaternary until the present, we evaluate the effect of paleoclimatic shifts on the present genetic structure and population differentiation of *Hypsiboas lundii*, a treefrog endemic to the South American Cerrado savanna. Our results show a recent and strong genetic divergence across the Cerrado landscape from west to east that does not seem congruent with any current physical barrier to gene flow. We found substantial support for the role of past climatic changes in population differentiation when controlling for the geographic distance among populations. Areas of more stable climatic conditions allowing population persistence since the Pleistocene seems to have played a strong role establishing the present genetic structure in this treefrog, a pattern consistent with isolation by instability, even though isolation by distance also played a part. Our study highlights the strong effects of Pleistocene climatic shifts also on tropical savannas.

INTRODUCTION

The climatic fluctuations of the Pleistocene glacial cycles have greatly impacted the distribution of organisms, imprinting on their populations a genetic signature of such changes (Hewitt 2000; Davis & Shaw 2001). Even though the genetic consequences of the ice ages are well documented for many organisms (including *Homo sapiens*) in the Northern Hemisphere (Hewitt 2000; Hewitt 2004), less evidence exists for the Southern Hemisphere, possibly due to the reduced impact of the glaciations in this region. The extent to which glacial cycles have impacted South America, for example, remains elusive (Werneck 2011). Although glaciers covered only small portions of South America, especially in the southern region and the Andes (Vuilleumier 1971), cycles of contraction and expansion of forests and savannas during glacial and interglacial periods are thought to have occurred (Van der Hammen 1974) and to have greatly affected vegetation structure and distribution of organisms (Bonaccorso *et al*. 2006; Saia *et al*. 2008).

The effect of Pleistocene climatic fluctuations on South America has always been a controversial topic, especially since it was first used as an explanation for the high number of species in the Amazon (Haffer 1969). The refuge hypothesis (Haffer 1969) predicted that glacial periods promoted fragmentation of forest into isolated patches serving as fauna refuges, which then spurred allopatric speciation. However, since additional evidence did not support fragmentation of the Amazon forest (Colinvaux *et al*. 2000), elevated speciation rates in the Pleistocene (Rull 2008) or refugia locations (Nelson *et al*. 1990), the impact of climatic fluctuations during the last glaciations has been strongly dismissed for tropical regions (Willis & Whittaker 2000). Recently, however, historical processes involving recent past climatic changes have been evoked to explain the strong genetic structure found in the Brazilian Atlantic Forest (Carnaval *et al*.

2009). The prediction of historical stable habitats since the Pleistocene also seems to be an important factor in explaining patterns of endemic diversity (Carnaval *et al*. 2014), species richness (Graham *et al*. 2006; Werneck *et al*. 2012b), and genetic structure in several other tropical regions (Carnaval & Bates 2007; Bell *et al*. 2010; He *et al*. 2013), supporting that substantial range shifts associated with glacial cycles occurred in the Tropics. Moreover, tropical species in particular might have been affected by temperature shifts during glaciations due to their narrow range of thermal tolerance (Deutsch *et al*. 2008).

The spatial genetic structure of populations reflects both historical and current ecological factors limiting gene flow (Epps & Keyghobadi 2015). But, historical factors such as paleoclimate dynamics or past barriers to gene flow that are not currently present (e.g. paleodrainages, paleorivers) seem to have a long and lasting contribution to the current pattern of genetic differentiation. In many cases, those can better explain the population genetic variation across the landscape than current climate or current barriers to gene flow (He *et al*. 2013; Nascimento *et al*. 2013; Ortego *et al*. 2014; Thomaz *et al*. 2015). Therefore, current low levels of admixture may not overwrite the genetic signatures of demographic and distributional changes associated with past environmental conditions. In fact, historical demographic processes have to be carefully considered when trying to find current ecological correlates to gene flow (Dyer *et al*. 2010).

In addition, understanding how historical factors influence population structure and gene flow can also add important insights about drivers of diversification and their relative contribution to high biodiversity regions. For example, the Cerrado, which is the largest and most species-rich tropical savanna and a global hotspot of biodiversity (Myers *et al*. 2000; Klink & Machado 2005), is remarkable for high endemicity levels reported for plants (44%; Klink & Machado 2005), frogs (53%; Valdujo *et al*. 2012), and

squamate reptiles (39%; Nogueira *et al*. 2011). Yet, the evolutionary and ecological processes contributing to this high species diversity are still not very well comprehended (Silva & Bates 2002; Werneck 2011). This savanna is part of an open dry diagonal in South America separating the Amazon forest in the northwest and the Atlantic forest in the southeast, and composed of a mosaic of heterogeneous vegetation ranging from grasslands to gallery forest along rivers. Adaptation to fire seems to have been an important and recent promoter for the wood plant radiation in this habitat (Simon *et al*. 2009), and geological activities, in particular the uplift of the Central Brazilian Shield, have been regarded as more important for population differentiation in vertebrates (Werneck *et al*. 2012a; Domingos *et al*. 2014; Santos *et al*. 2014). And, although some studies have associated intraspecific levels of differentiation to past climatic changes (Ramos *et al*. 2007; Novaes *et al*. 2010; Prado *et al*. 2012; Diniz-Filho *et al*. 2015), an explicit test demonstrating the significant association explaining the current genetic structure was only conducted for an endemic tree species in the Cerrado (Diniz-Filho *et al*. 2015). In contrast, very limited influence of paleoclimate has been reported for Cerrado lizards (Werneck *et al*. 2012a; Santos *et al*. 2014).

Available paleodistribution models based on paleoclimate for the Cerrado savanna predicts a dynamic history of vegetation instability over the late Quaternary, including both Pleistocene and Holocene epochs (Werneck *et al*. 2012b). According to Werneck *et al*. (2012b), the Cerrado was somewhat widespread over the Last Interglacial period (LIG) 120k years ago, characterized by a climate that was 0.5 to 2ºC warmer and slightly drier than present. Its distribution was greatly contracted during the Last Glacial Maxima (LGM) 21k years ago, which was 1 to 3.5ºC colder and drier than present. And, more recently, during the mid-Holocene 6k years ago, climate within the Cerrado was very similar to the present day conditions, with a distribution not substantially different from

present distribution. Although this dynamic response contrasts with the expected expansion of savanna towards adjacent forested regions during the dry, cold LGM (Ledru 2002), it highlights the dynamic climate and vegetation changes in the Cerrado throughout the late Quaternary.

Herein, we studied the phylogeography and population divergence of a widespread treefrog species that inhabits the gallery forests within the Cerrado savanna to understand the promoters of shallow diversification in this savanna. For this purpose, we used genomic-wide SNPs and a mitochondrial marker to obtain a complete picture of the genetic and geographic structure and population differentiation in this species. To evaluate whether environmentally stable areas throughout climatic fluctuations of the late Quaternary could explain the observed spatial structure for the treefrog *Hypsiboas lundii*, we used species distribution models for the present and predictions to three other time periods in the past to generate a stability surface predicting areas where populations have persisted. This surface was used to generate stability resistance distances among populations to be correlated with their genetic distances, an approach we denominate here as "isolation by instability". We then used different correlation approaches to test this hypothesis, controlling for the effect of geographic distance on population differentiation, a widespread phenomenon known as isolation by distance (Wright 1943).

MATERIALS AND METHODS

Geographic sampling and DNA extraction

We sampled 214 individuals of *Hypsiboas lundii* from 47 widespread localities in the Cerrado region (Table 3.1), in addition to two individuals of the sister species *Hypsiboas pardalis* from two different localities in the Atlantic Forest (Table S1).

Geographic sampling encompassed the entire known distribution for *Hypsiboas lundii* (Fig. 3.1). We isolated genomic DNA from liver or muscle samples preserved in 99% ethanol using a DNeasy Blood & Tissue Kit (Quiagen Inc., CA). The quality of our DNA extraction was visualized on a 1% agarose gel and quantified using a Quibit® 2.0 fluorometer (Life Technologies, NY).

ddRAD sequencing and assembly (SNP dataset)

We used the ddRADseq method to obtain single nucleotide polymorphisms (SNP data) in short loci distributed across in the entire genome. Following the protocol of Peterson *et al*. (2012), we prepared genomic libraries for all individuals of *Hypsiboas lundii* (214) and the two outgroup individuals of *H. pardalis*. In summary, two enzymes SphI and MspI (New England Biolabs, MA) were used to digest 250 ng of freshly extracted DNA for each sample. Each DNA digest was purified with Agencourt AMPure beads (Beckman Coulter, CA), then ligated with a pair of customized Illumina adaptors, which included a 5 bp unique barcode in the P1 adaptor and a 10 bp degenerate base region (DBR) in the P2 adaptor following Schweyen *et al*. (2014). The inclusion of a DBR (5' NNNNNIIICC 3') in the P2 adaptor allowed us to mitigate PCR duplication bias during library preparation (see our bioinformatics pipeline), one of the main criticisms of ddRADseq protocol regarding bias in genotype calls (Puritz *et al*. 2014). After ligation, equimolar quantities of purified samples were pooled into libraries of up to 48 samples. Fragment size selection followed using BluePippin (Sage Science, MA) with a size range of 448-496 bp. Libraries were purified with Dynabeads® M-270 Streptavidin (Life Technologies, NY) and amplified with a 12-cycle PCR using Phusion® High Fidelity polymerase kit (New England Biolabs, MA) and indexed primers provided in Peterson *et al*. (2012). The quality of each library was confirmed on a 2100 Bioanalyzer (Agilent

Technologies, CA) before sequencing on an Illumina® Hiseq 2500 PE (2 x 125bp) at the Genomic Sequencing and Analysis Facility (GSAF) at the University of Texas at Austin. We used a full sequencing lane for every two libraries of 48 samples each. Raw sequencing reads are available at NCBI SRA.

The bioinformatics pipeline was performed on HPC clusters at Texas Advanced Computing Center (TACC) as follows. Libraries were first demultiplexed by their individual barcodes using deML (Renaud *et al*. 2015) based on the maximum likelihood of assignment with the default quality scores cutoff. We filtered out PCR duplicates eliminating identical reads (window of 100 bp after the DBR) with an identical DBR matching pattern (10 bp) for each individual using a custom python script available at https://github.com/marimira/PCRdups_remover. This step reduced the amount of over-represented fragments resulting from PCR duplication, which could lead to a skewed allele frequency. We then used pyRAD (Eaton 2014) for subsequent quality filter, *de novo* assembly using both paired-end reads, and genotype calling. After filtering out low quality reads, discarding reads with more than four low quality bases (score <20), reads were clustered within and across individuals with a 90% similarity threshold using VSEARCH (https://github.com/torognes/vsearch). During the genotype-calling step, a consensus sequence is generated for all loci within individuals taking into account an estimated sequencing error rate (ϵ) and heterozygosity (π) across all sites (ϵ=0.00164 and π=0.00776 for this study). Loci with low coverage (< 6X) within individuals were excluded. A final alignment step of each loci across samples was then performed using MUSCLE (Edgar 2004), which allowed for indels.

Our final loci consisted of at least 229 bp after removing barcodes, enzyme cut-sites, DBR, and after concatenating both paired-end read with gaps allowed. As a filter for potential paralogs, loci with more than 2 alleles per individual (*H. lundii* is diploid)

are excluded, in addition to individual consensus sequences with more than 10 heterozygote sites, and final loci containing >20 SNPs. We did not set a filter based on shared heterozygote sites across samples in pyRAD. Because we are interested in the population genetics of a single species, setting stringent filters for shared heterozygotes across individuals could potentially erase real genetic signatures of demographic processes; therefore we set this value to the maximum number of 214 individuals. Finally, to explore the effect of missing data (due to allele dropout or low coverage for some loci), five different datasets with differing levels of loci coverage across samples were produced in pyRAD. We output matrices of loci and SNPs shared among at least 40%, 50%, 60%, 70%, and 80% of all *Hypsiboas lundii* individuals (Table 3.1). In addition, matrices were produced either including or excluding the outgroup individuals from the assembly, to generate input datasets for different kinds of analyses that either allowed establishing samples as outgroups (e.g. phylogenetic analyses) or not (e.g. population genetic analyses).

Population structure and SNP phylogeography

To infer population genetic structure in *H. lundii,* we used STRUCTURE v.2.3.4 (Pritchard *et al*. 2000), a Bayesian clustering method in which we implemented an admixture model with correlated allele frequency among populations (Falush *et al*. 2003). For this analysis, we used matrices of unlinked SNPs (one per locus of paired-end read) not including the outgroup individuals. We conducted preliminary runs with different levels of missing data and found very congruent results among them. However, with very low levels of missing data, the number of SNPs is substantially reduced, which could affect the statistical power to evaluate the number of clusters. Therefore, here and in following analyses, we used the matrix with up to 50% missing data as a good trade-off

between the total number of SNPs and missing data allowed per locus, besides the additional benefit of including some fast evolving sites that could add in identifying the fine genetic structure. We performed 10 runs for each K populations ranging from 1-8 and a MCMC run of 400,000 steps after a burn-in of 100,000 steps. We used a fixed lambda of 0.28 in all runs, which was first inferred from runs with K=1, as recommended in the software manual for SNP datasets. To allow for parallel runs on a computer cluster, we used the R package ParallelStructure (Besnier & Glover 2013). The appropriate number of genetic clusters was evaluated by examining the mean and variance of the log-likelihood for each K and implementing the ΔK statistic of Evanno *et al*. (2005) in STRUCTURE HARVESTER (Earl & vonHoldt 2011). Graphs were produced summarizing all runs across K values in the CLUMPAK server (Kopelman *et al*. 2015).

To infer phylogeographic structure we used two approaches using the SNP dataset. First, to infer the relationship among all individuals including the outgroup, we used a maximum likelihood approach in RAxML v.8 (Stamatakis 2014) using the conditional likelihood method (Lewis 2001) as an acquisition bias correction for dealing only with variable sites. We used the K80 model of nucleotide substitution without rate heterogeneity (ASC_GTRCAT) for all five concatenated SNP datasets with different levels of missing data. The datasets were pre-processed using the phrynomics R package (Leaché *et al*. 2015) to filter out invariable and non-binary sites. Branch support was estimated using bootstrap with an automatic stopping criterion once enough replicates had been generated (Pattengale *et al*. 2010). Then, for a species tree analysis of the populations we used SVDquartets v.1.0 (Chifman & Kubatko 2014), a recent coalescent method for SNP data that evaluates quartets of taxa, implemented in PAUP* v.4.0a147 (Swofford 2002). We used 10 million random quartets (13.5% of all possible quartets) with 100 bootstraps, setting *Hypsiboas pardalis* as the outgroup.

Mitochondrial DNA sequencing and analyses

To confirm the phylogeographic structure for *Hypsiboas lundii* and check the congruence among nuclear and mitochondrial markers, we also sequenced the ND2 mitochondrial gene (1032 bp) for a subset of 63 individuals (Table S1) encompassing all 47 localities (1-2 ind./locality), in addition to two samples of the outgroup *H. pardalis*. We amplified this region with TaKaRa Ex Taq® polymerase kit (Clontech laboratories Inc., CA) using the primers ND2F (5' aag acc tcc ttg ata ggg ag 3') and ND2R (5' tgc tta ggg ctt tga agg cyc 3'), and the following thermocycler conditions: initial denaturation at 94ºC for 2 min, followed by 35 cycles (30 sec denaturation at 94ºC, 30 sec annealing at 48ºC, 1 min extension at 72ºC), and a final extension at of 7 min at 72ºC. PCR products were purified using Genelute PCR clean-up (Sigma-Aldrich, MO) and sequenced at both directions at Macrogen Inc. (Seoul, South Korea). We used Geneious® 7.1.7 (Kearse *et al*. 2012) to edit and assemble individual sequences, and to perform a multiple alignment across all individuals using the MUSCLE plug-in (Edgar 2004). Sequences are available on GenBank.

PartitionFinder v. 1.1.1 (Lanfear *et al*. 2012) was used to obtain the following best-fit partitioning scheme and substitution models using BIC and linked branch lengths: (1) HKY+I for ND2 position1, (2) HKY for ND2 position2 + tRNAMet, (3) TrN+G for ND2 position 3. To generate a mitochondrial phylogeographic hypothesis for *Hypsiboas lundii* and estimate divergence times, we used BEAST v.1.7.5 (Drummond *et al*. 2012) with the partition scheme above. We used the uncorrelated lognormal relaxed clock with the published ND2 substitution rate of 1.3% per lineage per million year obtained from another frog species (Macey *et al*. 2001). We ran BEAST under the coalescent prior of constant population size for 10^7 generations sampled at every 10^3 generations. Convergence and stationarity (ESS > 200) were assessed in Tracer v.1.6 with a burn-in of

20%, and the maximum clade credibility tree was summarized using TreeAnnotator. In addition to the calibrated tree, we also used BEAST v.1.7.5 to reconstruct demographic history in *H. lundii* using Bayesian skyline plots (Drummond 2005) including all populations and also for each of the recovered lineages, using the same settings as above.

Species distribution models (present and past predictions)

We modeled *Hypsiboas lundii* distribution under the current climate and predicted to three other time periods in the past: the mid-Holocene (6kya), the Last Glacial Maximum - LGM (21kya), and the Last Interglacial period - LIG (120kya) to identify areas of historical persistence of the species under past climate change. Current and past climatic data were downloaded from the WorldClim database (Hijmans *et al*. 2005) at a spatial resolution of 30'' (current and LIG) or 2.5' (LGM and mid-Holocene). Data files at 2.5' were downscaled to 30'' with bilinear interpolation prior to analyses. The LGM and mid-Holocene climatic data were derived from the MPI-ESM-P General Circulation Model developed by Max-Plank Institute, and the LIG data from Otto-Bliesner *et al*. (2006).

All climatic layers were cropped to the same extent (0º to 30ºS and 35ºW to 65ºW). We used the variance inflation factor (VIF) to detect collinearity problems among climatic predictors using the usdm R package (Naimi *et al*. 2013). Predictors strongly correlated with one or more predictors have large VIF values adding collinearity problems to the model. Predictors with large VIF values were then removed in a stepwise procedure until all remaining predictors had VIF < 10. The nine climatic predictors remaining in the final model were: mean diurnal temperature range (BIO2), isothermality (BIO3), mean temperature of wettest quarter (BIO8), mean temperature of warmest quarter (BIO 10), precipitation of wettest month (BIO13), precipitation of driest month

(BIO14), precipitation seasonality (BIO15), precipitation of warmest quarter (BIO18), precipitation of coldest quarter (BIO19).

To model potential distributions, we used the random forests machine learning algorithm (Breiman 2001), implemented in the randomForest R package (Liaw & Wiener 2002), using the classification model with 500 trees and 3 variables sampled at each split. Occurrence data (74 points) included our field collection GPS records and geo-referenced records from herpetological collections in Brazil with the species identification verified for each locality to minimize occurrence errors introduced in the model. Because of the lack of true absence points, 100 pseudo-absence points were randomly generated in a reduced spatial extent of about 5º around the outmost presence points (8ºS to 27ºS, 60ºW to 38ºW). We followed the recommendations of Barbet-Massin *et al*. (2012), using the same number of pseudo-absence points as for presence points, averaging the model predictions of 10 runs with different sets of random points, and using a buffer excluding the immediate area around each presence point. However, for our study, we used a buffer of 40km instead, which was more appropriate due to our reduced spatial extent.

To evaluate the model performance under current climate, we applied a 5-fold cross-validation method, in which presence and pseudo-absence data points are randomly split into 5 subsamples and each one is once applied to test the model trained with the remaining 80% of the data. We used the area under the curve (AUC) of the receiver operating characteristic (ROC) plot, to evaluate commission errors (specificity) and omission errors (sensitivity). The AUC is a threshold-independent measure of the ability of the model to discriminate between presence and absence sites, and values significantly higher than 0.5 indicate models better than random. After ensuring good model performance and prediction for current climate, each model replicate was then projected onto each of the three past climates (Mid-Holocene, LGM and LIG). Presence/absence

maps are the final output of random forests classification models, however, after averaging 10 runs, we considered presence only when predicted in most of the replicate runs (threshold ≥ 0.6). Finally, to evaluate range stability through time, the presence/absence map outputs for all four time periods were overlaid and checked for areas of overlap among all time periods, generating a climatic stability surface with values ranging from 0 (never present) to 4 (present during all four times).

Testing the effect of historical climate change on gene flow and population structure

To test the hypothesis that present genetic structure of *Hypsiboas lundii* was shaped primarily by distributional shifts caused by climatic changes in the Quaternary, we examined if our climatic stability surface could explain genetic differentiation among populations. We used the stability surface as a conductance map in Circuitscape v.4.0 to generate pair-wise resistance distances among populations. Therefore, areas predicting lower stability throughout time have lower conductance values possessing very high resistance to dispersal, a model we refer here as isolation by instability (IBI). The resistance distances are calculated using a connectivity model based on circuit theory using an eight-neighbor cell scheme to estimate an average resistance value among populations in a heterogeneous landscape (McRae *et al*. 2008). Pairwise F_{ST} was used as a proxy of gene flow among populations. We calculated pairwise F_{ST} (Weir & Cockerham 1984) in VCFtools (Danecek *et al*. 2011) for the 23 populations containing at least four individuals, since F_{ST} can be precisely estimated for sample sizes as small as four for large numbers of SNPs (>1000) (Willing *et al*. 2012). We then tested our model, which is a variation of the isolation by resistance model (McRae 2006), using a Mantel test between the pair-wise genetic and the resistance distance matrices, using also a partial Mantel test to control for geographic distance among populations. Mantel and partial Mantel tests

were conducted in the ecodist R package (Goslee & Urban 2007) with 1000 randomizations and 500 bootstraps for significance test. We also compared the IBI to the null model of isolation by distance (IBD) (Wright 1943) using a matrix of pair-wise Euclidean geographic distances instead, which was calculated in the R package raster. Lastly, we evaluated IBI using the multiple regression analysis on distance matrices proposed by Legendre *et al*. (1994) in ecodist R package. We used this approach to confirm the results of the partial mantel test, for which criticism has been raised concerning its lower statistical power (Legendre & Fortin 2010).

RESULTS

ddRADseq processing

We aimed for ~2 million reads per sample, however, after the steps of demultiplexing, PCR duplicate removal, and filtering out low quality reads, we obtained an average of ~1.6 million reads per sample. Our pyRAD pipeline generated an average of 33,597 loci per sample with average depth coverage of ~18X, for all loci above the coverage threshold of 6X which passed the paralog filters. After aligning all loci across samples, our ddRADseq data matrices varied substantially between 169 to 14,397 loci depending on the amount of missing data allowed at any locus. Moreover, including the outgroup in the assembly negatively impacted our paralog filters, since the final matrices including the outgroup had less loci than excluding it from the assembly (Table 3.2).

Population structure and SNP phylogeography

Our STRUCTURE analyses recovered four main genetic clusters for *Hypsiboas lundii* in the Brazilian Cerrado arranged from west to east as follows: the Western cluster comprising localities 1 to 3 from Mato Grosso state; the Central-Western cluster

comprising localities 4 to 7 from Mato Grosso and western Goiás state; the Central cluster including localities 8 to 22 from Goiás state; the Southeastern cluster including localities 23 to 47, mostly from Minas Gerais and São Paulo state (Figs. 3.1 and 3.2). We observed an appropriate level of convergence across replicate runs given the small variance around the mean log-likelihood across K (Fig 3.2). A steady increase in the mean likelihood is observed until K=4, which according to the ΔK statistic of Evanno *et al*. (2005) is the most likely number of clusters (Fig 3.2), then it levels off, showing substantially lower improvement in the model with larger K values.

Except for minor differences, the topologies generated in RAxML for SNP data with different levels of missing data were very similar recovering the same geographic clades and the four genetic clusters with very high support (Fig 3.3A). The only exception was the topology generated with up to 20% missing data loci, which did not recover the Central cluster as a clade and also depicted a slightly different relationship among the clusters, with decreased bootstrap support overall (Supp. Info.). Thus, settings that severely limit the amount of missing data in final SNP datasets (particularly when using a *de-novo* assembly with no reference genome), might pose a challenge to recovering enough SNPs with strong phylogenetic signal for reliable inferences.

The population tree generated in SVDquartets also recovered the four geographic clusters with very high support (Fig 3.3B). The major differences between the phylogenetic trees and the STRUCTURE plot were: the assignment of individuals from locality 13 (domin), which is assigned to the Central cluster by STRUCTURE, but the sister clade to the Central and Southeast clades in all phylogenies; and locality 4 (garca), which is assigned to the Central-West cluster by STRUCTURE, but part of the West clade in the SNP phylogenies (although part of the Central-West clade in the mitochondria). Low levels of current or ancestral gene flow might explain differences between STRUCTURE

and tree analyses, given that in the STRUCTURE plot individuals of those localities show evidence for admixture with a small posterior probability of assignment to a different cluster.

Mitochondrial phylogeography

We also recovered strong geographic structure in the mitochondria for *Hypsiboas lundii*, portraying the same west to east geographic pattern. Our mitochondrial topology showed an overall agreement with the SNP topologies, except for the relationships involving the Central-West cluster, which were not consistent with the clades in the mitochondrial tree. The conflicts between mitochondrial and SNP phylogenies included three localities of the Central-West cluster (altog, serra, caiap) and locality 13 (domin), which is part of the Central clade (with low support), in agreement with the STRUCTURE assignment (Fig 3.3C). The most recent common ancestor (MRCA) of *H. lundii* was estimated to be 2.1 My old, therefore, the diversification in this species is very recent occurring in the Pleistocene (< 2.6 Mya). Yet, our divergence time estimates show that all four recognized geographical clusters were already established by the LIG (120kya) and could have been impacted in a similar way by past climatic oscillations. The Western clade is the most basal clade in *H. lundii*, and is also the furthest from the sister species *H. pardalis* that is endemic to the Atlantic Forest, east from the Cerrado savanna. Despite the recent origination of *H. lundii* and the recent divergence of *H. pardalis* populations, the MRCA between *H. lundii* and *H. pardalis* was estimated to be 8.0 Mya.

The Bayesian skyline plot including all populations of *Hypsiboas lundii* showed that it experienced a significant increase in population size at around 100kya, coinciding with the last glaciations cycle of the Pleistocene, and a recent decline (Fig 3.4A). However, this decline could be an artifact for the presence of population structure using

the all the demes in the analysis (Heller *et al*. 2013). The Bayesian skyline plots for the Central and Southeast lineages confirmed the increasing trend, showing population expansion for both lineages (Figs. 3.4B and 3.4C). However, the slow steady increase in population size in the Central lineage contrasts with the recent and rapid increase in the Southeast lineage. Bayesian inference of population size could not be performed for the West and the Central-West lineages due to their small sample sizes.

Distribution shifts during glaciations (present and past predictions)

Most of the predicted distribution of *Hypsiboas lundii* for the current climate was within the Cerrado savanna; however the model overpredicted its distribution to the east, in the adjacent areas of the Atlantic Forest. The inclusion of several presence points in the contact zone between the Atlantic Forest and the Cerrado may explain this overprediction. The current climate model performed well with an average AUC of 0.81 for the bootstrapped test data (Table S2). The most important climatic predictors to the current model were the precipitation of wettest month (BIO13) and the mean temperature of warmest quarter (BIO 10) (Table S2). All of the past climate predictions departed considerably, showing a substantial range shift for *Hypsiboas lundii* across the late Quaternary. The mid-Holocene (6kya) prediction was the most similar to the current climate prediction. Conversely, the LGM and LIG predictions differed greatly from each other and from current climate prediction. The LGM predicted the largest area of occupancy, not currently occupied by the species (Fig 3.5), indicating an unexpected increase in habitat suitability, especially towards the northwest, during the dry and cold LGM. The LIG predicted the smallest area suitability, indicating that the LIG had a stronger effect constraining the distribution of *H. lundii* compared to the glacial period of the LGM. After obtaining a stability surface of all time periods, we recognized four main

areas of stability in which the species has been present since 120k years ago (Fig 3.6). These areas greatly overlap with the core distribution area of the four genetic clusters.

Testing the effect of historical climate change on population structure

We found a positive significant relationship between genetic and stability resistance distances (r^2 = 0.71, p = 0.001), corroborating the effect of isolation by instability (resistance to the stability surface) in the genetic differentiation of *Hypsiboas lundii* (Fig 3.6). However, we also found a positive significant relationship between genetic and geographic distances among populations (r^2 = 0.74, p = 0.001), corroborating also the effect of isolation by distance on the genetic structure (Fig 3.6). Due to the high correlation between the stability resistances and geographic distances (r^2 = 0.71, p = 0.001), we also performed statistical tests to evaluate the isolation by instability given the geographic distance among the populations. Using a partial Mantel test, we found that stability resistances were positively correlated with genetic distances, even when removing the effect of the geographic distance among populations (r^2 = 0.4, p = 0.03). Using a multiple regression on distance matrices, we found that geographic distance was the most important variable in the multiple regression model (ß < 0.001, p = 0.004), however, stability resistance distances also contributed significantly to the model (ß = 0.432, p = 0.04). The final regression model included both geographic and stability resistance effects on the genetic structure of *Hypsiboas lundii*.

DISCUSSION

We recovered a strong spatial genetic structure in *Hypsiboas lundii*, a treefrog endemic to the Cerrado savanna, with four well-delimited genetic clusters, which also comprised well-supported clades with the exception of the Central-West cluster. This

genetic structure was corroborated by genome-wide SNPs and mitochondrial DNA, showing a general consistency among markers and demonstrating a well-established genetic structure across the landscape. This structure was found to be associated to recent past climate changes that affected population persistence in climatically unstable areas. Our study is not the first to detect a significant role of past climatic fluctuation structuring populations in the Cerrado. Diniz-Filho *et al*. (2015) found genetic differentiation (F_{ST}) and low levels of heterozygosity significantly associated with shifts in climatic suitability since the LGM for *Eugenia dysenterica*, an endemic Cerrado tree. Other studies in tropical or sub-tropical areas have also recently found a significant association between genetic structure and climatic stable areas over the last glaciations (Carnaval & Bates 2007; Bell *et al*. 2010; Devitt *et al*. 2013; Gugger *et al*. 2013; He *et al*. 2013; Ortego *et al*. 2014). Therefore, the model we refer as isolation by instability over the late Quaternary might be prevalent in places mildly affected by glaciations promoting population differentiation.

Our model of isolation by instability (IBI) was supported even when taking into account the geographic distance among populations. Nonetheless, isolation by distance (IBD) was also found to be a strong determinant of population differentiation in this species, which was somewhat expected given the widespread distribution of this treefrog and the usual phylopatry exhibited by anurans (Wells 2007). In fact, IBD seems ubiquitous among organisms (Meirmans 2012). A pattern of IBD can, in many instances, produce misleading hierarchical genetic clusters, with the identification of well-delimited clusters even when none is existent (Meirmans 2012) especially with biased sampling across the landscape (Schwartz & McKelvey 2008). Regardless of strong evidence for IBD, our study included comprehensive population sampling across the species distribution, and found strongly supported genetic structure using various methods,

including Bayesian clustering and phylogenetic analyses. Moreover, STRUCTURE performs better than any other clustering algorithm to detect clinal variation when present (Chen *et al*. 2007), thus we are confident the clusters identified are not spurious clinal gradient. And, the incongruent pattern found for the Central-West cluster, for example, was also recognized by STRUCTURE based on the admixed ancestry coefficient for some localities in this cluster.

The incongruent relationship of the Central-West populations in the mitochondrial and SNP phylogenies may have resulted from recent or ancestral levels of gene flow among populations of this cluster and the closest populations from other clusters or some shared ancestral polymorphism (Muir & Schlötterer 2004). Mito–nuclear biogeographic discordance is widespread among animals and most commonly the result of geographic isolation with varied levels of mitochondrial introgression after secondary contact (Toews & Brelsford 2012). Mito–nuclear conflicts have been reported for several other frog species that show a geographic pattern that is consistent with recent or ancestral mitochondrial introgression (Toews & Brelsford 2012; and references therein; Bryson & Smith 2014). This discordance might be facilitated by the smaller effective population size of the haploid maternally inherited mitochondria, for which even low levels of gene flow might be sufficient to establish mitochondrial introgression more rapidly than nuclear markers producing the biogeographic conflict (Chan & Levin 2005). Recent or ancestral mitochondrial introgression might explain populations in the Central-West cluster grouping with different clades in the mitochondrial tree compared to the SNP phylogenies. However, this pattern was only observed in the Central-West cluster, which is located in a more scattered stability area than the other clusters.

A drastic distributional shift was inferred in *Hypsiboas lundii* throughout the late Quaternary. Its predicted widespread range during the LGM was not expected, given that

the Cerrado distribution was predicted to have contracted during this time (Werneck *et al*. 2012b). Aside from all sources of uncertainty in species distribution modeling, when applied to an entire ecosystem, as in the case of the Cerrado (Werneck *et al*. 2012b), it might be problematic because responses to climate change might be species-specific and the entire ecosystem might not respond as a single unit (Collevatti *et al*. 2012). This is especially relevant for the Cerrado, which is a heterogeneous complex ecoregion of different vegetation formations ranging from open grasslands to gallery forests; therefore different responses to climatic changes are expected for species inhabiting different habitats with different environmental requirements. In addition support to our climatic modeling, our Bayesian skyline plots indicated of a pronounced population expansion after the LIG (120kya) towards the LGM (21kya), comprising another independent evidence for the range expansion in the LGM. Interestingly, a phylogeographic study of another Cerrado treefrog also corroborates our findings reporting evidence for population expansion during the Pleistocene (at 89kya) in the same areas: the Central and the Southeast (Prado *et al*. 2012).

Although the cold and dry LGM was initially proposed to have restricted the distribution of forest species in South America, a recent study has challenged this traditional view. Leite *et al*. (2016), showed the predicted distribution for several Atlantic Forest species to be displaced north (warmer) towards lower elevation and less topographically complex areas, allowing a more widespread distribution than the present or the LIG. Our results support this trend also for Cerrado species, at least the ones inhabiting mesic habitats of gallery forests within this savanna. The distribution of *H. lundii* in the LGM expanded north and west in the Cerrado, towards less topographic complex landscape. In contrast, the LIG was the period predicted to have mostly constrained the distribution of this treefrog, playing a greater role in structuring

populations across the landscape that was still maintained with very limited levels of gene flow. However, the complete geographical separation of the Central-West from the remaining clusters by unsuitable conditions did not happen until the mid-Holocene, which may have contributed to the conflicting pattern in the Central-West cluster.

The geographic structure for *Hypsiboas lundii* the does not seem congruent with any current discrete barrier to gene flow, such as topographic barriers (large rivers or river basins) that separate the geographic clusters in this savanna. In fact, when compared to other South American ecoregions such as the Amazon (with its large river barriers), the Atlantic Forest (with a steep latitudinal gradient), or the Andes (with great topographic complexity and elevation range), the Cerrado seems to be a barrier-free habitat. In addition, some other studies investigating the population structure of plants, frogs, and lizards in the Cerrado have found a congruent pattern of stronger genetic differentiation in the west-east or northwest-southeast axis (Ramos *et al*. 2007; Ramos *et al*. 2008; Prado *et al*. 2012; Santos *et al*. 2014; Barbosa *et al*. 2015; Guarnizo *et al*. 2016), without recognizing any conspicuous geographical barrier. This supports our view on the main role of distribution shifts during climatic oscillations, leading to the recent genetic differentiation in this savanna. The west-east pattern suggests that historical climatic suitability might have forced distribution shifts around this main axis, which was important for delimiting the current genetic structure for many species.

Despite being distributed across a broad elevational range, most of the stable areas identified encompass major plateaus and more elevated areas within the treefrog distribution. Previous studies have also identified climatic stable areas in the Cerrado plateaus (Werneck *et al*. 2012b; Santos *et al*. 2014). Climate can interact synergistically with topography, producing a complex gradient set of conditions that might play an important role buffering climatic fluctuations and allowing persistence of populations

over time. In addition, the importance of plateaus for Cerrado biodiversity has already been pointed out by studies showing that areas of higher endemism in the Cerrado are also usually concentrated on plateaus (Nogueira *et al*. 2011; Valdujo *et al*. 2012).

CONCLUSION

The impact of climatic fluctuations of the Quaternary in tropical regions has been a controversial issue over the last decades. Our study adds to the large body of evidence showing that climate has been unstable over the Quaternary in the tropical areas of South America. This instability promoted distributional shifts of animals and plants, as demonstrated here for a treefrog species, which greatly affected the genetic differentiation of its populations and gene flow across the landscape, a model we refer here as "isolation by instability." A pattern of isolation by distance was also found to be significant in our study, but isolation by instability was still significant when removing the effect of geographic distance among populations.

The Cerrado is an important global hotspot of biodiversity in South America but only recently the evolutionary and ecological drivers of diversification in this savanna have been investigated (Silva & Bates 2002; Simon *et al*. 2009; Werneck 2011). Here, we demonstrate that past climatic fluctuations might be important in driving intraspecific genetic structure in this habitat as has already been demonstrated for the Atlantic Forest (Carnaval *et al*. 2009). Genetic differentiation in *Hypsiboas lundii* was significantly correlated with the climatic instability distance among populations, a metric based on the persistence of populations across time using past predictions of species distribution models over the late Quaternary. Understanding how species in biodiverse regions responded to past climate changes can be relevant considering the current scenario of

climate change and global warming. The conditions under which populations survived climate change and habitat instability might give us perspective on how species will cope with current and future levels of climate change and habitat fragmentation.

Table 3.1: Sampled localities of *Hypsiboas lundii* in the Brazilian Cerrado with geographic coordinates in decimal degrees.

#	Code	Municipality	State	N	Long	Lat	Cluster
1	guima	Chapada dos Guimarães	MT	10	-55.804	-15.472	West
2	jacia	Jaciara	MT	9	-55.036	-15.983	West
3	rondo	Rondonópolis	MT	1	-54.736	-16.669	West
4	altog	Alto Garças	MT	5	-53.480	-17.043	CentralWest
5	garca	Barra do Garças	MT	6	-52.254	-15.873	CentralWest
6	serra	Serranópolis	GO	9	-51.963	-18.288	CentralWest
7	caiap	Caiapônia	GO	4	-51.833	-16.963	CentralWest
8	goias	Goiás	GO	11	-50.115	-16.001	Central
9	campe	Campestre de Goiás	GO	2	-49.697	-16.737	Central
10	petro	Petrolina de Goiás	GO	2	-49.215	-16.256	Central
11	limpo	Campo Limpo de Goiás	GO	4	-49.082	-16.292	Central
12	piren	Pirenópolis	GO	5	-48.964	-15.850	Central
13	domin	São Domingos	GO	3	-46.326	-13.395	Central
14	caval	Cavalcante	GO	3	-47.432	-13.812	Central
15	teres	Teresina de Goiás	GO	7	-47.262	-13.875	Central
16	parai	Alto Paraíso de Goiás	GO	4	-47.506	-14.138	Central
17	alian	São João d'Aliança	GO	13	-47.508	-14.649	Central
18	brasi	Brasília	DF	1	-47.857	-15.677	Central
19	desco	Santo Antônio do Descoberto	GO	3	-48.272	-16.097	Central
20	luzia	Luziânia	GO	6	-48.176	-16.299	Central
21	migue	São Miguel do Passa Quatro	GO	1	-48.546	-17.042	Central
22	pires	Pires do Rio	GO	1	-48.258	-17.237	Central
23	catal	Catalão	GO	9	-47.728	-17.923	Southeast
24	unaii	Unaí	MG	9	-47.259	-16.408	Southeast
25	parac	Paracatu	MG	5	-46.824	-17.149	Southeast
26	gauch	Chapada Gaúcha	MG	7	-45.543	-15.364	Southeast
27	santa	Santa Fé de Minas	MG	11	-45.414	-16.673	Southeast
28	janua	Januária	MG	10	-44.240	-15.109	Southeast
29	pinhe	João Pinheiro	MG	10	-46.185	-17.677	Southeast
30	pocoe	Claro dos Poções	MG	5	-44.140	-17.104	Southeast
31	crist	Cristália	MG	7	-42.873	-16.635	Southeast
32	riach	Santana do Riacho	MG	3	-43.721	-19.168	Southeast
33	horiz	Belo Horizonte	MG	1	-43.912	-19.946	Southeast
34	curve	Curvelo	MG	1	-44.606	-19.152	Southeast
35	furna	São José da Barra	MG	2	-46.327	-20.687	Southeast
36	roque	São Roque de Minas	MG	3	-46.622	-20.262	Southeast

Table 3.1: continued.

37	araxa	Araxá	MG	1	-46.941	-19.594	Southeast
38	perdi	Perdizes	MG	8	-47.140	-19.209	Southeast
39	sacra	Sacramento	MG	2	-47.293	-19.875	Southeast
40	pedre	Pedregulho	SP	1	-47.463	-20.244	Southeast
41	simao	São Simão	SP	1	-47.607	-21.497	Southeast
42	srita	Santa Rita do Passa Quatro	SP	1	-47.591	-21.726	Southeast
43	claro	Rio Claro	SP	1	-47.699	-22.313	Southeast
44	cosmo	Cosmópolis	SP	1	-47.219	-22.646	Southeast
45	botuc	Botucatu	SP	1	-48.445	-22.886	Southeast
46	bauru	Bauru	SP	1	-49.072	-22.243	Southeast
47	assis	Assis	SP	3	-50.375	-22.598	Southeast

Table 3.2: Output of pyRAD for number of loci and SNPs according to the maximum amount of missing data allowed per locus, either including the outgroup samples or not.

Missing data	**Min. ind.**	**Loci**	**Unlinked SNPs**	**Total SNPs**	**Outgroup**
60%	86	13895	13837	148665	yes
50%	107	7489	7460	81909	yes
40%	128	3330	3319	36917	yes
30%	150	823	820	9243	yes
20%	171	169	168	1823	yes
60%	86	14397	14314	144363	no
50%	107	7776	7734	78716	no
40%	128	3473	3453	35278	no
30%	150	859	851	8909	no
20%	171	180	177	1775	no

Min. ind. = minimum number of ingroup individuals set to share a locus.
Loci = total paired-end reads in the final dataset (including non-variable).
Unlinked SNPs = one single SNP taken from each variable loci.
Total SNPs = all variable sites across all variable loci.

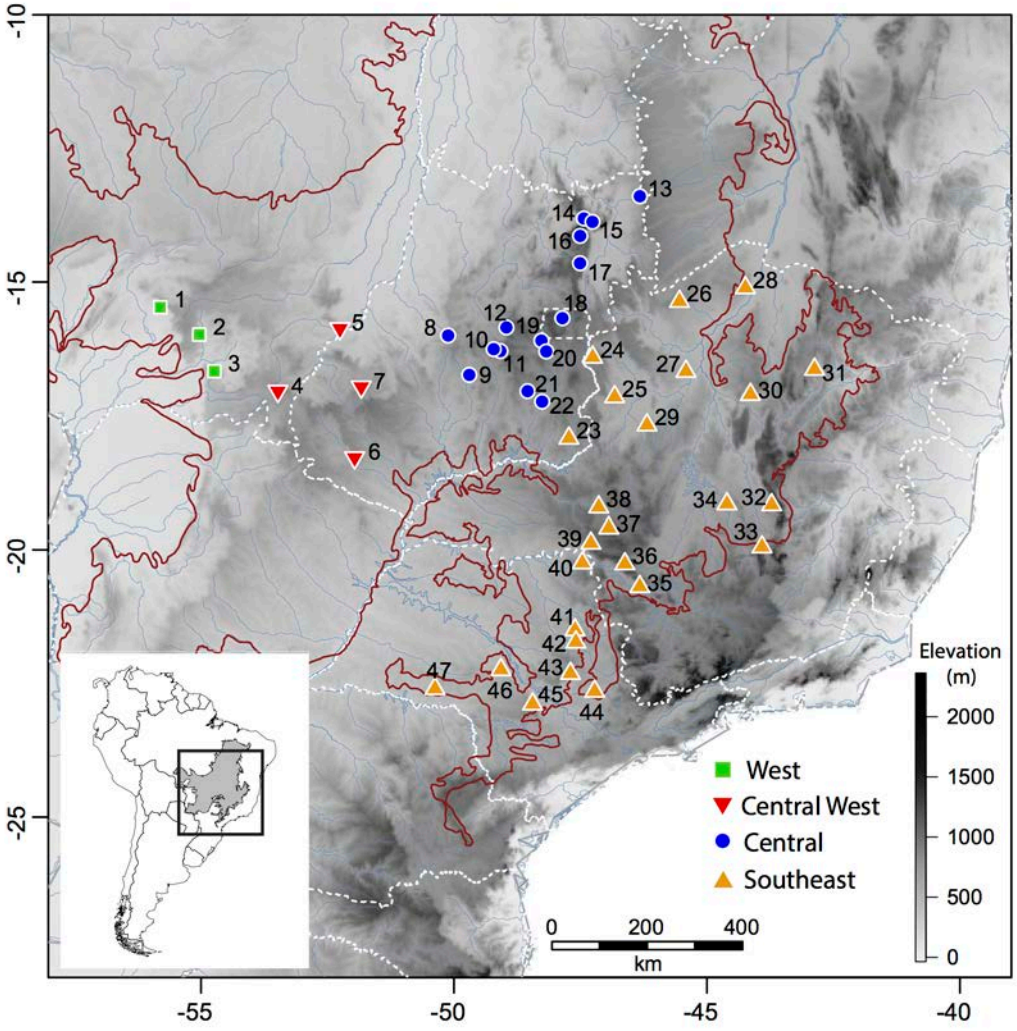

Figure 3.1: Distribution of sampled localities for *Hypsiboas lundii* in the Cerrado savanna of South America identified by different symbols that correspond to the genetic clusters found in this study (West, Central-West, Central, Southeast). Numbers correspond to each of the 47 localities in Table 1. Cerrado limits are shown in both maps.

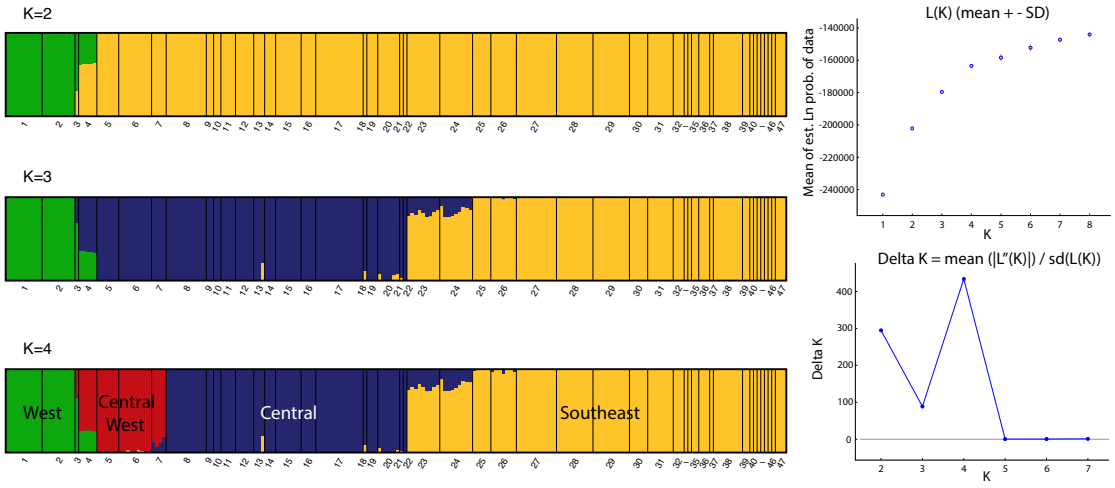

Figure 3.2: Results from STRUCTURE using 50% missing data SNP matrix. In the left, the individual assignment plots for K=2 to 4, and the in the right, the plots showing the mean log probability of the data L(K) and the ΔK statistic of Evanno *et al*. (2005) showing a higher value for K=4.

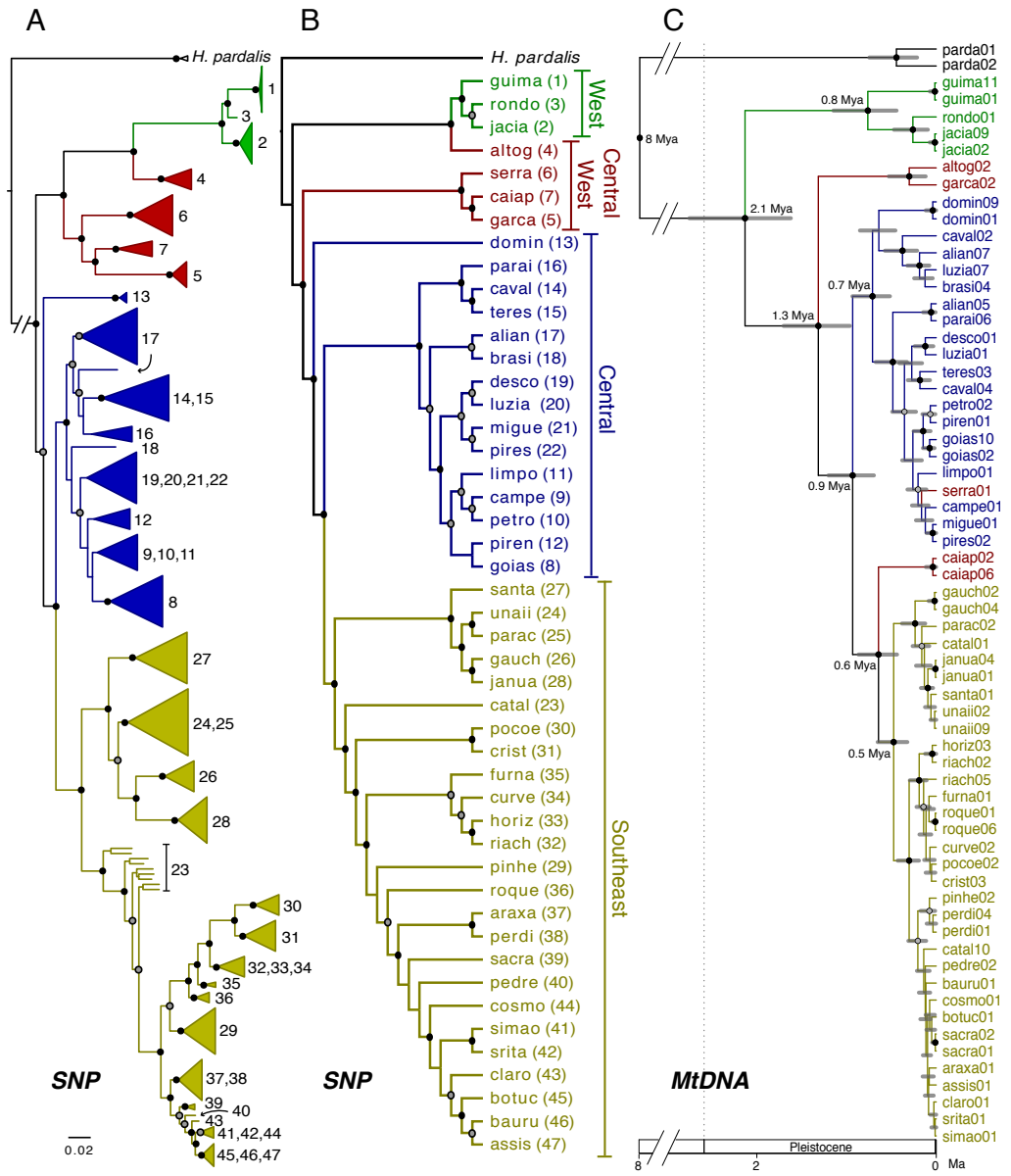

Figure 3.3: Phylogenetic trees colored according to genetic clusters. Node support >95% as black dots, and >70% shown as gray dots. **A.** Maximum-likelihood tree of all individuals inferred in RAxML using SNP matrix with up to 50% missing data. Branch lengths are scaled to nucleotide substitutions, but branch leading to outgroup was shortened to facilitate visualization. **B.** Population tree inferred in SVDquartets using SNP matrix with up to 50% missing data. Cladogram depicts no branch length information. **C.** Chronogram inferred in BEAST using the ND2 mitochondrial marker. Branch lengths are scaled to the time axis in the bottom and the 95% CI of divergence times are shown as gray bars at the nodes.

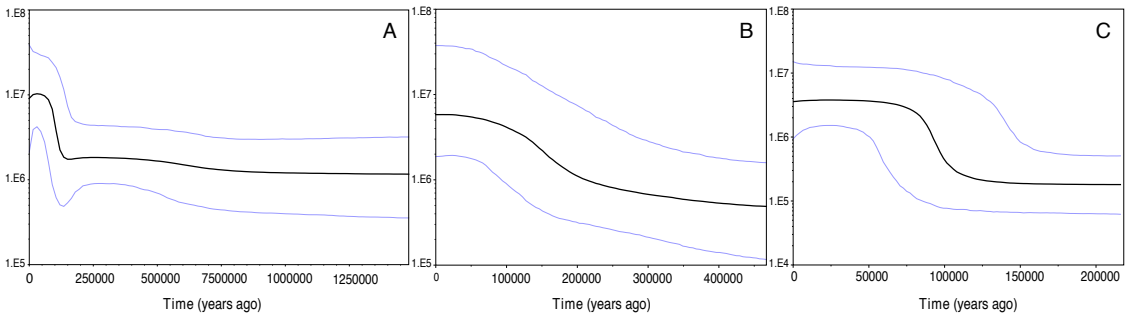

Figure 3.4: Bayesian skyline plots (BSP) showing variation in effective population sizes through time. **A.** For all populations. **B.** For the Central lineage. **C.** For the Southeast lineage. The dark thinker line represents the median population size, and the thinner lines, the 95% posterior probabilities.

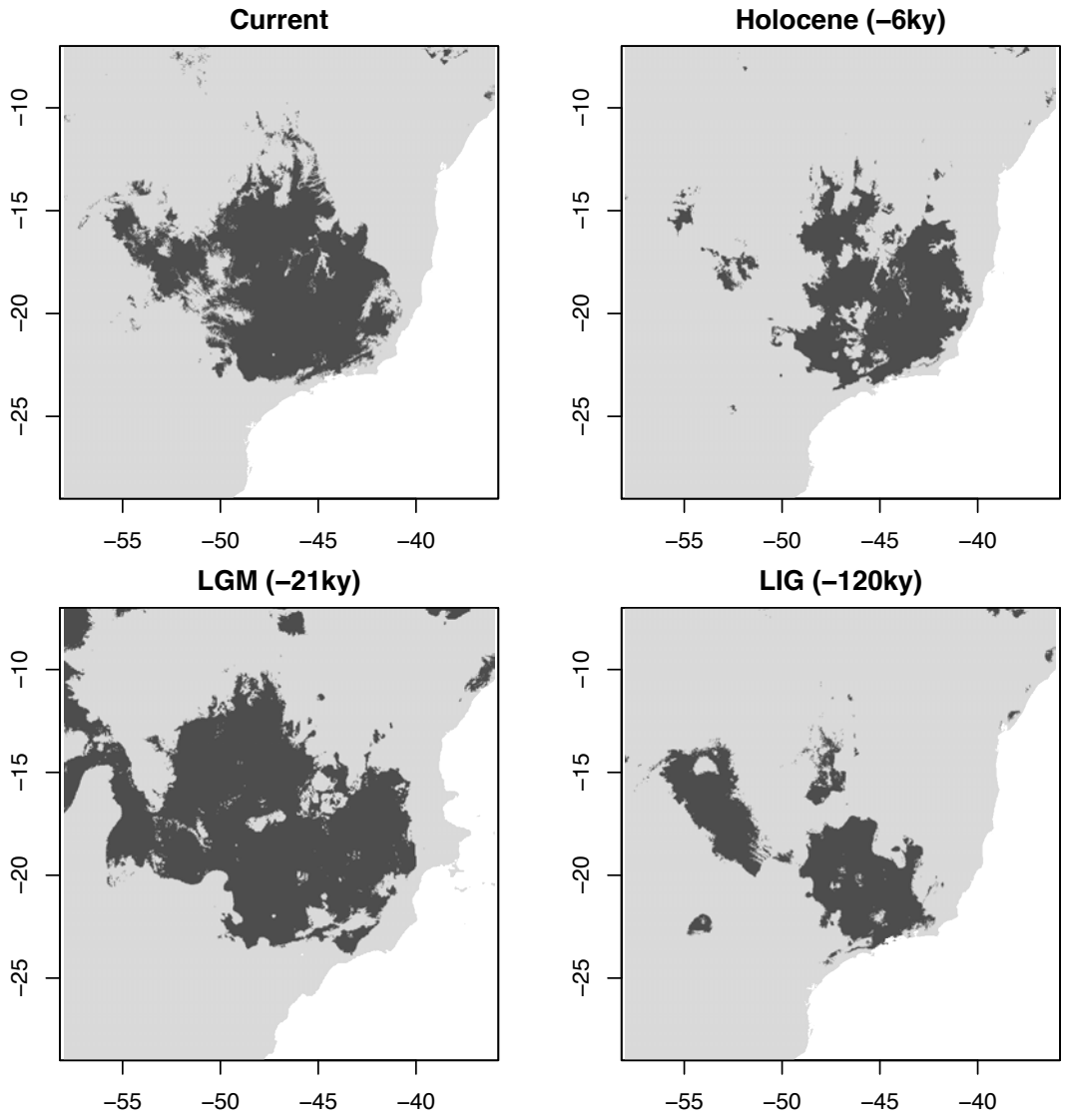

Figure 3.5: Species distribution models for *Hypsiboas lundii* using random forest under the current climate, and predicted to three different times in the past: the mid-Holocene, the Last Glacial Maxima (LGM), and the Last Interglacial Maxima (LIG).

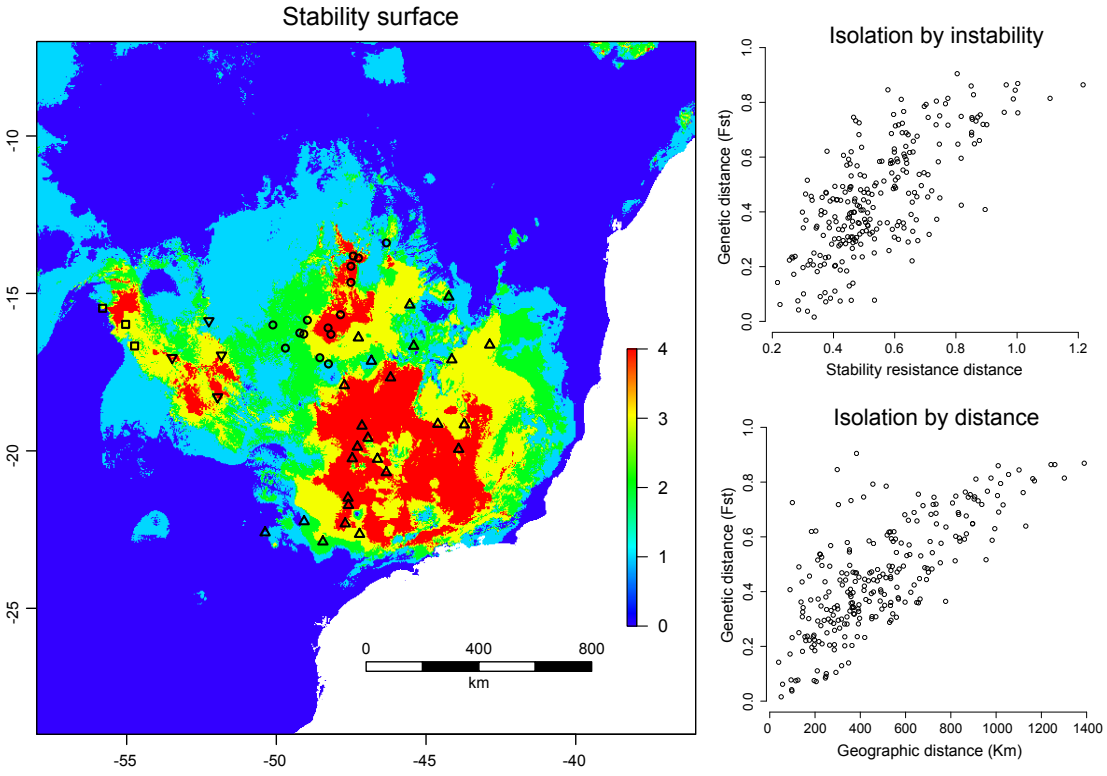

Figure 3.6: Stability surface generated by overlaying the four species distribution models in Fig. 3.5 ranging from 0 (never present) to 4 (always present). Higher values mean higher stability and population persistence over time. Populations are represented by symbols corresponding to the four genetic clusters indentified. Note the correspondence of the four main stable areas at the core distribution of the four genetic clusters. Plots on the right side show correlations between genetic distance and stability resistance (isolation by instability) and between genetic and geographic distance (isolation by distance).

CONCLUSION

The Cerrado is an important global hotspot of biodiversity in South America but only recently the evolutionary and ecological drivers of diversification in this savanna have been investigated (Silva & Bates 2002; Simon *et al*. 2009; Werneck 2011). The present study greatly contributes to the understanding of the Cerrado diversification adding to the increasing number of recent studies addressing diversity patterns in this rich savanna.

The overall large scope of this dissertation approaching diversification at different time scales allows some general interpretation of the species-level biogeographic patterns based on population-level divergences. In this study, I was able to connect how ecological gradients and climatic fluctuations can promote divergence of populations across the landscape. I argue that the establishment of a strong population structure associated with pockets of stable climate in the Cerrado Neotropical savanna (Chapter 3) may be the primary promoter for recent cryptic allopatric speciation events in the absence of substantial morphological or sexual behavior divergence (Chapter 2), implying that there is no evident selective pressure for increased species differentiation other than differential adaptation to distinct local climates. This increased response to the impact of recent climatic changes may ultimately lead to the biogeographic pattern found of recent endemic and cryptic lineages inside the Cerrado savanna (Chapter 1). Future studies focusing on genes of climatic adaptation importance for treefrogs related to important traits conferring resistance to desiccation or temperature regulation might be able to detect genetic signatures for selection across the landscape, especially in lineages widely distributed in the Cerrado.

References

Antonelli A, Verola CF, Parisod C, Gustafsson ALS (2010) Climate cooling promoted the expansion and radiation of a threatened group of South American orchids (Epidendroideae: Laeliinae). *Biological Journal of the Linnean Society* **100**, 597-607.

Barbet-Massin M, Jiguet F, Albert CH, Thuiller W (2012) Selecting pseudo-absences for species distribution models: how, where and how many? *Methods in Ecology and Evolution* **3**, 327-338.

Barbosa ACOF, Collevatti RG, Chaves LJ, *et al*. (2015) Range-wide genetic differentiation of Eugenia dysenterica (Myrtaceae) populations in Brazilian Cerrado. *Biochemical Systematics and Ecology* **59**, 288-296.

Beebee TJC (1995) Amphibian breeding and climate. *Nature* **374**, 219-220.

Beheregaray LB (2008) Twenty years of phylogeography: the state of the field and the challenges for the Southern Hemisphere. *Molecular Ecology* **17**, 3754-3774.

Bell RC, Parra JL, Tonione M, *et al*. (2010) Patterns of persistence and isolation indicate resilience to climate change in montane rainforest lizards. *Molecular Ecology* **19**, 2531-2544.

Berner D (2011) Size correction in biology: how reliable are approaches based on (common) principal component analysis? *Oecologia* **166**, 961-971.

Besnier F, Glover KA (2013) ParallelStructure: A R Package to Distribute Parallel Runs of the Population Genetics Program STRUCTURE on Multi-Core Computers. *Plos One* **8**, e70651.

Bonaccorso E, Koch I, Peterson AT (2006) Pleistocene fragmentation of Amazon species' ranges. *Diversity and Distributions* **12**, 157-164.

Bouckaert R, Heled J, Kühnert D, *et al*. (2014) BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS computational biology* **10**, e1003537.

Bouckaert RR (2010) DensiTree: making sense of sets of phylogenetic trees. *Bioinformatics* **26**, 1372-1373.

Boul KE, Chris Funk W, Darst CR, Cannatella DC, Ryan MJ (2007) Sexual selection drives speciation in an Amazonian frog. *Proceedings of the Royal Society B-Biological Sciences* **274**, 399-406.

Breiman L (2001) Random Forests. *Machine Learning* **45**, 5-32.

Bremer K (1992) Ancestral areas: a cladistic reinterpretation of the center of origin concept. *Systematic Biology* **41**, 436-445.

Brown JM, Hedtke SM, Lemmon AR, Lemmon EM (2010) When trees grow too long: Investigating the causes of highly inaccurate Bayesian branch-length estimates. *Systematic Biology* **59**, 145-161.

Brown JM, Lemmon AR (2007) The importance of data partitioning and the utility of bayes factors in bayesian phylogenetics. *Systematic Biology* **56**, 643-655.

Bryson RW, Chaves J, Smith BT, *et al*. (2014) Diversification across the New World within the 'blue' cardinalids (Aves: Cardinalidae). *Journal of Biogeography* **41**, 587-599.

Bryson RW, Smith BT (2014) The role of mitochondrial introgression in illuminating the evolutionary history of Nearctic treefrogs. *Zoological Journal of the Linnean Society* **172**, 103–116.

Burnaby TP (1966) Growth-invariant discriminant functions and generalized distances. *Biometrics* **22**, 96-110.

Burnham KP, Anderson DR (2002) *Model Selection and Multi-Model Inference: A Practical Information Theoretic Approach*, 2nd ed edn. Springer Verlag, Berlin.

Camargo A, Morando M, Avila LJ, Sites J, Jack W (2012) Species delimitation with ABC and other coalescent-based methods: a test of accuracy with simulations and an emprirical example with lizards of the *Liolaemus darwinii* complex (Squamata: Liolaemidae). *Evolution* **66**, 2834-2849.

Caramaschi U, Cruz CAG (2000) Duas espécies novas de *Hyla* Laurenti, 1768 do Estado de Goiás, Brasil (Amphibia, Anura, Hylidae). *Boletim do Museu Nacional, Nova Série Zoológica*, 1-12.

Caramaschi U, Cruz CAG (2004) Duas novas espécies de *Hyla* do grupo de *H. polytaenia* Cope, 1870 do Sudeste do Brasil (Amphibia, Anura, Hylidae). *Arquivos do Museu Nacional, Rio de Janeiro* **62**, 247-254.

Carnaval AC, Bates JM (2007) Amphibian DNA shows marked genetic structure and tracks Pleistocene climate change in northeastern Brazil. *Evolution* **61**, 2942-2957.

Carnaval AC, Hickerson MJ, Haddad CFB, Rodrigues MT, Moritz C (2009) Stability predicts genetic diversity in the Brazilian Atlantic Forest hotspot. *Science* **323**, 785-789.

Carnaval AC, Moritz C (2008) Historical climate modelling predicts patterns of current biodiversity in the Brazilian Atlantic forest. *Journal of Biogeography* **35**, 1187-1201.

Carnaval AC, Waltari E, Rodrigues MT, *et al*. (2014) Prediction of phylogeographic endemism in an environmentally complex biome. *Proceedings of the Royal Society B-Biological Sciences* **281**, 20141461-20141461.

Carstens BC, Pelletier TA, Reid NM, Satler JD (2013) How to fail at species delimitation. *Molecular Ecology* **22**, 4369-4383.

Castroviejo-Fisher S, Guayasamin JM, Gonzalez-Voyer A, Vilà C (2014) Neotropical diversification seen through glassfrogs. *Journal of Biogeography* **41**, 66-80.

Caviedes-Solis IW, Bouzid NM, Banburry BL (2015) Uprooting phylogenetic uncertainty in coalescent species delimitation: A meta-analysis of empirical studies. *Curr Zool*.

Chan KMA, Levin SA (2005) Leaky prezygotic isolation and porous genomes: rapid introgression of maternally inherited DNA. *Evolution* **59**, 720-729.

Chen C, Durand E, Forbes F, François O (2007) Bayesian clustering algorithms ascertaining spatial population structure: a new computer program and a comparison study. *Molecular Ecology Notes* **7**, 747-756.

Chifman J, Kubatko L (2014) Quartet inference from SNP data under the coalescent model. *Bioinformatics* **30**, 3317-3324.

Colinvaux PA, De Oliveira PE, Bush MB (2000) Amazonian and neotropical plant communities on glacial time-scales: The failure of the aridity and refuge hypotheses. *Quaternary Science Reviews* **19**, 141-169.

Collevatti RG, Terribile LC, de Oliveira G, *et al*. (2012) Drawbacks to palaeodistribution modelling: the case of South American seasonally dry forests. *Journal of Biogeography* **40**, 345-358.

Colli GR (2005) As origens e a diversificação da herpetofauna do Cerrado. In: *Cerrado: Ecologia, Biodiversidade e Conservação* (eds. Scariot A, Sousa-Silva JC, Felfili JM), pp. 247-264. Ministério do Meio Ambiente, Brasília, DF.

Condamine FL, Sperling FAH, Wahlberg N, Rasplus JY, Kergoat GJ (2012) What causes latitudinal gradients in species diversity? Evolutionary processes and ecological constraints on swallowtail biodiversity. *Ecology Letters* **15**, 267-277.

Cope ED (1869) Seventh contribution to the herpetology of tropical America. *Proceedings of the American Philosophical Society* **11**, 147-192.

Cruz CAG, Caramaschi U (1998) Definição, composição e distribuição geográfica do grupo de *Hyla polytaenia* Cope, 1870 (Amphibia, Anura, Hylidae). *Boletim do Museu Nacional, Nova Série Zoológica* **392**, 1-19.

Cusimano N, Renner SS (2010) Slowdowns in diversification rates from real phylogenies may not be real. *Systematic Biology* **59**, 458-464.

Danecek P, Auton A, Abecasis G, *et al*. (2011) The variant call format and VCFtools. *Bioinformatics* **27**, 2156-2158.

Darst CR, Cannatella DC (2004) Novel relationships among hyloid frogs inferred from 12S and 16S mitochondrial DNA sequences. *Molecular Phylogenetics and Evolution* **31**, 462-475.

Davis MB, Shaw RG (2001) Range shifts and adaptive responses to Quaternary climate change. *Science* **292**, 673-679.

de Queiroz K (1998) The general lineage concept of species, species criteria, and the process of specieation: a conceptual unification and terminological recommendations. In: *Endless forms: species and speciation* (eds. Howard DJ, Berlocher SH). Oxford University Press, Oxford.

de Queiroz K (2007) Species concepts and species delimitation. *Systematic Biology* **56**, 879–886.

Derryberry EP, Claramunt S, Derryberry G, *et al*. (2011) Lineage diversification and morphological evolution in a large-scale continental radiation: the Neotropical ovenbirds and woodcreepers (Aves:Furnariidae). *Evolution* **65**, 2973-2986.

Deutsch CA, Tewksbury JJ, Huey RB, *et al*. (2008) Impacts of climate warming on terrestrial ectotherms across latitude. *Proceedings of the National Academy of Sciences, USA* **105**, 6668-6672.

Devitt TJ, Devitt SEC, Hollingsworth BD, McGuire JA, Moritz C (2013) Montane refugia predict population genetic structure in the Large-blotched *Ensatina* salamander. *Molecular Ecology* **22**, 1650-1665.

Diniz-Filho JAF, Barbosa ACOF, Collevatti RG, *et al*. (2015) Spatial autocorrelation analysis and ecological niche modelling allows inference of range dynamics driving the population genetic structure of a Neotropical savanna tree. *Journal of Biogeography* **43**, 167–177.

Diniz-Filho JAF, Loyola RD, Raia P, Mooers AO, Bini LM (2013) Darwinian shortfalls in biodiversity conservation. *Trends In Ecology & Evolution* **28**, 689-695.

Domingos FMCB, Bosque RJ, Cassimiro J, *et al*. (2014) Out of the deep: cryptic speciation in a Neotropical gecko (Squamata, Phyllodactylidae) revealed by species delimitation methods. *Molecular Phylogenetics and Evolution* **80**, 113-124.

Drummond AJ (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology and Evolution* **22**, 1185-1192.

Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* **29**, 1969-1973.

Duellman WE (2001) *The Hylid Frogs of Middle America* Society for the study of Amphibians and Reptiles, Ithaca, NY.

Duellman WE, De La Riva I, Wild ER (1997) Frogs of the *Hyla armata* and *Hyla pulchella* groups in the Andes of South America, with definitions and analyses of

phylogenetic relationships of Andean groups of *Hyla*. *Scientific Papers (Natural History Museum The University of Kansas)* **3**, 1-41.

Dyer RJ, Nason JD, Garrick RC (2010) Landscape modelling of gene flow: improved power using conditional genetic distance derived from the topology of population networks. *Molecular Ecology* **19**, 3746-3759.

Earl DA, vonHoldt BM (2011) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* **4**, 359-361.

Eaton DAR (2014) PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics* **30**, 1844-1849.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**, 1792-1797.

Edwards AWF (1992) *Likelihood* Jonh Hopkins University Press, Baltimore.

Edwards DL, Knowles LL (2014) Species detection and individual assignment in species delimitation: can integrative data increase efficacy? *Proceedings of the Royal Society B-Biological Sciences* **281**, 20132765-20132765.

Edwards EJ, Osborne CP, Strömberg CAE, *et al*. (2010) The origins of C_4 grasslands: Integrating evolutionary and ecosystem science. *Science* **328**, 587-591.

Epps CW, Keyghobadi N (2015) Landscape genetics in a changing world: disentangling historical and contemporary influences and inferring change. *Molecular Ecology* **24**, 6021-6040.

Etienne RS, Haegeman B, Stadler T, *et al*. (2011) Diversity-dependence brings molecular phylogenies closer to agreement with the fossil record. *Proceedings of the Royal Society of London B: Biological Sciences* **279**, 1300-1309.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology* **14**, 2611-2620.

Evans BJ, Kelley DB, Tinsley RC, Melnick DJ, Cannatella DC (2004) A mitochondrial DNA phylogeny of African clawed frogs: phylogeography and implications for polyploid evolution. *Molecular Phylogenetics and Evolution* **33**, 197-213.

Faivovich J, Garcia PCA, Ananias F, *et al*. (2004) A molecular perspective on the phylogeny of the *Hyla pulchella* species group (Anura, Hylidae). *Molecular Phylogenetics and Evolution* **32**, 938-950.

Faivovich J, Haddad CFB, Garcia PCA, *et al*. (2005) Systematic review of the frog family Hylidae, with special reference to Hylinae: phylogenetic analysis and taxonomic revision. *Bulletin of the American Museum of Natural History* **294**, 1-240.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**, 1567-1587.

Filzmoser P, Hron K, Reimann C (2012) Interpretation of multivariate outliers for compositional data. *Computers and Geosciences* **39**, 77-85.

Fine PVA, Zapata F, Daly DC (2014) Investigating processes of Neotropical rain forest tree diversification by examining the evolution and historical biogeography of the Protieae (Burseraceae). *Evolution* **68**, 1988–2004.

Flot J-F (2010) Seqphase: a web tool for interconverting phase input/output files and fasta sequence alignments. *Molecular Ecology Resources* **10**, 162-166.

Flower BP, Kennett JP (1994) The middle Miocene climatic transition: East Antarctic ice sheet development, deep ocean circulation and global carbon cycling. *Palaeogeography, Palaeoclimatology, Palaeoecology* **108**, 537-555.

Fouquet A, Gilles A, Vences M, *et al*. (2007) Underestimation of species richness in Neotropical frogs revealed by mtDNA analyses. *Plos One* **2**, e1109.

Fujita MK, Leaché AD, Burbrink FT, McGuire JA, Moritz C (2012) Coalescent-based species delimitation in an integrative taxonomy. *Trends in Ecology & Evolution* **27**, 480-488.

Funk WC, Caminer M, Ron SR (2012) High levels of cryptic species diversity uncovered in Amazonian frogs. *Proceedings of the Royal Society B: Biological Sciences* **279**, 1806-1814.

Garreaud RD (2009) The Andes climate and weather. *Advances in Geosciences* **22**, 3-11.

Garzione CN, Hoke GD, Libarkin JC, *et al*. (2008) Rise of the Andes. *Science* **320**, 1304-1307.

Gingras B, Boeckle M, Herbst CT, Fitch WT (2012) Call acoustics reflect body size across four clades of anurans. *Journal of Zoology* **289**, 143-150.

Goebel AM, Donnelly JM, Atz ME (1999) PCR primers and amplification methods for 12S ribosomal DNA, the control region, cytochrome oxidase I, and cytochrome b in bufonids and other frogs, and an overview of PCR primers which have amplified DNA in amphibians successfully. *Molecular Phylogenetics and Evolution* **11**, 163-199.

Goldberg EE, Roy K, Lande R, Jablonski D (2005) Diversity, endemism, and age distributions in macroevolutionary sources and sinks. *The American Naturalist* **165**, 623-633.

Goslee SC, Urban DL (2007) The ecodist Package for Dissimilarity-based Analysis of Ecological Data. *Journal of Statistical Software* **22**, 1-19.

Graham CH, Moritz C, Williams SE (2006) Habitat history improves prediction of biodiversity in rainforest fauna. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 632-636.

Graham CH, Ron SR, Santos JC, Schneider CJ, Moritz C (2004) Integrating phylogenetics and environmental niche models to explore speciation mechanisms in dendrobatid frogs. *Evolution* **58**, 1781-1793.

Guarnizo CE, Werneck FP, Giugliano LG, *et al*. (2016) Cryptic lineages and diversification of an endemic anole lizard (Squamata, Dactyloidae) of the Cerrado hotspot. *Molecular Phylogenetics and Evolution* **94**, 279-289.

Guerra MA, Ron SR (2008) Mate choice and courtship signal differentiation promotes speciation in an Amazonian frog. *Behavioral Ecology* **19**, 1128-1135.

Guerra MA, Ryan MJ, Cannatella DC (2014) Ontogeny of Sexual Dimorphism in the Larynx of the Túngara Frog, Physalaemus pustulosus. *Copeia* **2014**, 123-129.

Gugger PF, Ikegami M, Sork VL (2013) Influence of late Quaternary climate change on present patterns of genetic variation in valley oak, *Quercus lobata* Née. *Molecular Ecology* **22**, 3598-3612.

Guimarães TCS, Figueiredo GB, Mesquita DO, Vasconcellos MM (2011) Ecology of *Hypsiboas albopunctatus* (Spix, 1824) (Anura, Hylidae) in a Neotropical savanna. *Journal of Herpetology* **45**.

Haffer J (1969) Speciation in Amazonian forest birds. *Science* **165**, 131-137.

He Q, Edwards DL, Knowles LL (2013) Integrative testing of how environments from the past to the present shape genetic structure across landscapes. *Evolution* **67**, 3386-3402.

Heinicke MP, Duellman WE, Trueb L, *et al*. (2009) A new frog family (Anura: Terrarana) from South America and an expanded direct-developing clade revealed by molecular phylogeny. *Zootaxa* **2211**, 1-35.

Heled J, Drummond AJ (2010) Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution* **27**, 570-580.

Heller R, Chikhi L, Siegismund HR (2013) The Confounding Effect of Population Structure on Bayesian Skyline Plot Inferences of Demographic History. *Plos One* **8**, e62992.

Hewitt G (2000) The genetic legacy of the Quaternary ice ages. *Nature* **405**, 907-913.

Hewitt GM (2004) Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society B: Biological Sciences* **359**, 183-195.

Hey J, Waples RS, Arnold ML, Butlin RK, Harrison RG (2003) Understanding and confronting species uncertainty in biology and conservation. *Trends in Ecology & Evolution* **18**, 597–603.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25**, 1965-1978.

Holman JA (2003) *Fossil frogs and toads of North America* Indiana University Press, Bloomington and Indianapolis.

Hoorn C, Wesselingh FP, ter Steege H, *et al*. (2010) Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. *Science* **330**, 927-931.

Hoskin CJ, James S, Grigg GC (2009) Ecology and taxonomy-driven deviations in the frog call-body size relationship across the diverse Australian frog fauna. *Journal of Zoology* **278**, 36-41.

Hughes CE, Pennington RT, Antonelli A (2013) Neotropical plant evolution: assembling the big picture. *Botanical Journal of the Linnean Society* **171**, 1-18.

Hutter CR, Guayasamin JM, Wiens JJ (2013) Explaining Andean megadiversity: the evolutionary and ecological causes of glassfrog elevational richness patterns. *Ecology Letters* **16**, 1135-1144.

IUCN (2014) The IUCN Red List of Threatened Species. http://www.iucnredlist.org.

Jablonski D (1993) The tropics as a source of evolutionary novelty through geological time. *Nature* **364**, 142-144.

Jablonski D, Roy K, Valentine JW (2006) Out of the tropics: Evolutionary dynamics of the latitudinal diversity gradient. *Science* **314**, 102-106.

Jolicoeur P (1963) The multivariate generalization of the allometry equation. *Biometrics* **19**, 497-499.

Kass RE, Raftery AE (1995) Bayes factors. *Journal of the American Statistical Association* **90**, 773-795.

Kearse M, Moir R, Wilson A, *et al*. (2012) Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647-1649.

Klink CA, Machado RB (2005) Conservation of the Brazilian Cerrado. *Conservation Biology* **19**, 707-713.

Köhler J, Koscinski D, Padial JM, *et al*. (2010) Systematics of Andean gladiator frogs of the *Hypsiboas pulchellus* species group (Anura, Hylidae). *Zoologica Scripta* **39**, 572–590.

Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I (2015) CLUMPAK: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources* **15**, 1179-1191.

Lanfear R, Calcott B, Ho SYW, Guindon S (2012) PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution* **29**, 1695-1701.

Latorre C, Quade J, McIntosh WC (1997) The expansion of C_4 grasses and global change in the late Miocene: Stable isotope evidence from the Americas. *Earth and Planetary Science Letters* **146**, 83-96.

Leaché AD, Banbury BL, Felsenstein J, de Oca An-M, Stamatakis A (2015) Short Tree, Long Tree, Right Tree, Wrong Tree: New Acquisition Bias Corrections for Inferring SNP Phylogenies. *Systematic Biology* **64**, 1032-1047.

Leaché AD, Fujita MK (2010) Bayesian species delimitation in West African forest geckos (*Hemidactylus fasciatus*). *Proceedings of the Royal Society B-Biological Sciences* **277**, 3071-3077.

Leaché AD, Fujita MK, Minin VN, Bouckaert RR (2014) Species delimitation using genome-wide SNP data. *Systematic Biology* **63**, 534-542.

Ledru M-P (2002) Late Quaternary Historyand Evolution of the Cerrados as Revealed by Palynological Records. In: *The Cerrados of Brazil: ecology and natural history of a Neotropical savanna* (eds. Oliveira PS, Marquis RJ), pp. 33-50. Columbia University Press, New York.

Legendre P, Fortin MJ (2010) Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. *Molecular Ecology Resources* **10**, 831-844.

Legendre P, Lapointe FJ, Casgrain P (1994) Modeling brain evolution from behavior: a permutational regression approach. *Evolution* **48**, 1487-1499.

Leite YLR, Costa LP, Loss AC, *et al*. (2016) Neotropical forest expansion during the last glacial period challenges refuge hypothesis. *Proceedings of the National Academy of Sciences, USA* **113**, 1008–1013.

Lemmon AR, Lemmon EM (2012) High-throughput identification of informative nuclear loci for shallow-scale phylogenetics and phylogeography. *Systematic Biology* **61**, 745-761.

Lemmon EM, Lemmon AR, Cannatella DC (2007) Geological and climatic forces driving speciation in the continentally distributed trilling chorus frogs (*Pseudacris*). *Evolution* **61**, 2086-2103.

Lewis PO (2001) A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology* **50**, 913-925.

Liaw A, Wiener M (2002) Classification and regression by randomForest. *R News* **2**, 18-22.

Liu K, Warnow TJ, Holder MT, *et al*. (2012) SATé-II: Very fast and accurate simultaneous estimation of multiple sequence alignments and phylogenetic trees. *Systematic Biology* **61**, 90-106.

Liu L, Pearl DK, Brumfield RT, Edwards SV (2008) Estimating species trees using multiple-allele DNA sequence data. *Evolution* **62**, 2080-2091.

Lutz B (1968) New Brazilian forms of *Hyla*. *Pearce-Sellards Series, Texas Memorial Museum* **10**, 1-18.

Mace GM (2004) The role of taxonomy in species conservation. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **359**, 711–719.

Macey JR, Strasburg JL, Brisson JA, *et al*. (2001) Molecular Phylogenetics of Western North American Frogs of the Rana boylii Species Group. *Molecular Phylogenetics and Evolution* **19**, 131-143.

Maciel NM, Collevatti RG, Colli GR, Schwartz EF (2010) Late Miocene diversification and phylogenetic relationships of the huge toads in the *Rhinella marina* (Linnaeus, 1758) species group (Anura: Bufonidae). *Molecular Phylogenetics and Evolution* **57**, 787-797.

Margules CR, Pressey RL (2000) Systematic conservation planning. *Nature* **405**, 243–253.

Marshall DC (2010) Cryptic failure of partitioned Bayesian phylogenetic analyses: lost in the land of long trees. *Systematic Biology* **59**, 108-117.

Martins FD (2011) Historical biogeography of the Brazilian Atlantic forest and the Carnaval-Moritz model of Pleistocene refugia: what do phylogeographical studies tell us? *Biological Journal of the Linnean Society* **104**, 499-509.

Matzke NJ (2013) BioGeoBEARS: Biogeography with bayesian (and likelihood) evolutionary analysis in R scripts, http://CRAN.R-project.org/package=BioGeoBEARS.

Matzke NJ (2014) Model selection in historical biogeography reveals that founder-event speciation is a crucial process in island clades. *Systematic Biology* **63**, 951-970.

McCormack JE, Hird SM, Zellmer AJ, Carstens BC, Brumfield RT (2013) Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution* **66**, 526-538.

McCoy MW, Bolker BM, Osenberg CW, Miner BG, Vonesh JR (2006) Size correction: comparing morphological traits among populations and environments. *Oecologia* **148**, 547-554.

McKenna DD, Farrell BD (2006) Tropical forests are both evolutionary cradles and museums of leaf beetle diversity. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 10947-10951.

McRae BH (2006) Isolation by resistance. *Evolution* **60**, 1551-1561.

McRae BH, Dickson BG, Keitt TH, Shah VB (2008) Using circuit theory to model connectivity in ecology, evolution, and conservation. *Ecology* **89**, 2712-2724.

Meirmans PG (2012) The trouble with isolation by distance. *Molecular Ecology* **21**, 2839-2846.

Menin M, Silva RA, Giaretta AA (2004) Reproductive biology of *Hyla goiana* (Anura, Hylidae). *Iheringia Série Zoologia*.

Mittelbach GG, Schemske DW, Cornell HV, *et al*. (2007) Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography. *Ecology Letters* **10**, 315-331.

Moen DS, Morlon H (2014) Why does diversification slow down? *Trends in Ecology & Evolution* **29**, 190-197.

Moen DS, Smith SA, Wiens JJ (2009) Community assembly through evolutionary diversification and dispersal in Middle American treefrogs. *Evolution* **63**, 3228-3247.

Mora C, Chittaro PM, Sale PF, Kritzer JP, Ludsin SA (2003) Patterns and processes in reef fish diversity. *Nature* **421**, 933-936.

Moreau CS, Bell CD (2013) Testing the museum versus cradle tropical biological diversity hypothesis: Phylogeny, diversification, and ancestral biogeographic range evolution of the ants. *Evolution* **67**, 2240-2257.

Moritz C (1994) Defining 'Evolutionarily Significant Units' for conservation. *Trends in Ecology & Evolution* **9**, 373–375.

Moritz C, Patton JL, Schneider CJ, Smith TB (2000) Diversification of rainforest faunas: An integrated molecular approach. *Annual Review of Ecology and Systematics* **31**, 533-563.

Muir G, Schlötterer C (2004) Evidence for shared ancestral polymorphism rather than recurrent gene flow at microsatellite loci differentiating two hybridizing oaks (*Quercus* spp.). *Molecular Ecology* **14**, 549-561.

Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Kent J (2000) Biodiversity hotspots for conservation priorities. *Nature* **403**, 853-858.

Naimi B, Hamm NAS, Groen TA, Skidmore AK, Toxopeus AG (2013) Where is positional uncertainty a problem for species distribution modelling? *Ecography* **37**, 191-203.

Nascimento FF, Lazar A, Menezes AN, *et al*. (2013) The role of historical barriers in the diversification processes in open vegetation formations during the Miocene/Pliocene using an ancient rodent lineage as a model. *Plos One* **8**, e61924.

Nelson BW, Ferreira CAC, da Silva MF, Kawasaki ML (1990) Endemism centres, refugia and botanical collection density in Brazilian Amazonia. *Nature* **345**, 714-716.

Niemiller ML, Graening GO, Fenolio DB, *et al*. (2013) Doomed before they are described? The need for conservation assessments of cryptic species complexes using an amblyopsid cavefish (Amblyopsidae: Typhlichthys) as a case study. *Biodiversity and Conservation* **22**, 1799-1820.

Nogueira C, Ribeiro S, Costa GC, Colli GR (2011) Vicariance and endemism in a Neotropical savanna hotspot: distribution patterns of Cerrado squamate reptiles. *Journal of Biogeography* **38**, 1907-1922.

Novaes RML, De Lemos JP, Ribeiro RA, Lovato MB (2010) Phylogeography of *Plathymenia reticulata* (Leguminosae) reveals patterns of recent range expansion towards northeastern Brazil and southern cerrados in eastern Tropical South America. *Molecular Ecology* **19**, 985-998.

Oliveira PS, Marquis RJ (2002) *The Cerrados of Brazil: Ecology and Natural History of a Neotropical Savanna* Columbia University Press, New York.

Oliver PM, Adams M, Doughty P (2010) Molecular evidence for ten species and Oligo-Miocene vicariance within a nominal Australian gecko species (*Crenadactylus ocellatus*, Diplodactylidae). *Bmc Evolutionary Biology* **10**, 386-386.

Oliver PM, Adams M, Lee MSY, Hutchinson MN, Doughty P (2009) Cryptic diversity in vertebrates: molecular data double estimates of species diversity in a radiation of Australian lizards (*Diplodactylus*, Gekkota). *Proceedings of the Royal Society B-Biological Sciences* **276**, 2001-2007.

Olson DM, Dinerstein E, Wikramanayake ED, *et al*. (2001) Terrestrial ecoregions of the world: A new map of life on Earth. *Bioscience* **51**, 933-938.

Ortego J, Gugger PF, Sork VL (2014) Climatically stable landscapes predict patterns of genetic structure and admixture in the Californian canyon live oak. *Journal of Biogeography* **42**, 328-338.

Otto-Bliesner BL, Marshall SJ, Overpeck JT, Miller GH, Hu A (2006) Simulating Arctic climate warmth and icefield retreat in the last interglaciation. *Science* **311**, 1751-1753.

Padial JM, Miralles A, de la Riva I, Vences M (2010) The integrative future of taxonomy. *Frontiers in Zoology* **7**, 16.

Paradis E, Claude J, Strimmer K (2004) APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**, 289-290.

Pattengale ND, Alipour M, Bininda-Emonds ORP, Moret BME, Stamatakis A (2010) How Many Bootstrap Replicates Are Necessary? *Journal of Computational Biology* **17**, 337-354.

Pepper M, Doughty P, Fujita MK, Moritz C, Keogh JS (2013) Speciation on the rocks: integrated systematics of the *Heteronotia spelea* species complex (Gekkota; Reptilia) from western and central Australia. *Plos One* **8**, e78110.

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. *Plos One* **7**, e37135.

Phillimore AB, Price TD (2008) Density-dependent cladogenesis in birds. *Plos Biology* **6**, 483-489.

Pigot AL, Phillimore AB, Owens IPF, Orme CDL (2010) The shape and temporal dynamics of phylogenetic trees arising from geographic speciation. *Systematic Biology* **59**, 660-673.

Potter PE (1997) The Mesozoic and Cenozoic paleodrainage of South America: a natural history. *Journal of South American Earth Sciences* **10**, 331-344.

Prado CPA, Haddad CFB, Zamudio KR (2012) Cryptic lineages and Pleistocene population expansion in a Brazilian Cerrado frog. *Molecular Ecology* **21**, 921-941.

Price SL, Powell S, Kronauer DJC, *et al*. (2014) Renewed diversification is associated with new ecological opportunity in the Neotropical turtle ants. *Journal of Evolutionary Biology* **27**, 242-258.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**, 945.

Puritz JB, Matz MV, Toonen RJ, *et al*. (2014) Demystifying the RAD fad. *Molecular Ecology* **23**, 5937-5942.

Pybus OG, Harvey PH (2000) Testing macro-evolutionary models using incomplete molecular phylogenies. *Proceedings of the Royal Society B-Biological Sciences* **267**, 2267-2272.

Pyron RA, Wiens JJ (2011) A large-scale phylogeny of Amphibia including over 2800 species, and a revised classification of extant frogs, salamanders, and caecilians. *Molecular Phylogenetics and Evolution* **61**, 543-583.

Pyron RA, Wiens JJ (2013) Large-scale phylogenetic analyses reveal the causes of high tropical amphibian diversity. *Proceedings of the Royal Society B-Biological Sciences* **280**, 20131622.

Quental TB, Marshall CR (2010) Diversity dynamics: molecular phylogenies need the fossil record. *Trends in Ecology & Evolution* **25**, 434-441.

Rabosky DL (2006) LASER: a maximum likelihood toolkit for detecting temporal shifts in diversification rates from molecular phylogenies. *Evolutionary Bioinformatics* **2**, 257-260.

Rabosky DL (2014) Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. *Plos One* **9**, e89543.

Rabosky DL, Grundler M, Anderson C, *et al*. (2014) BAMMtools: an R package for the analysis of evolutionary dynamics on phylogenetic trees. *Methods in Ecology and Evolution* **5**, 701-707.

Rabosky DL, Lovette IJ (2008) Density-dependent diversification in North American wood warblers. *Proceedings of the Royal Society B-Biological Sciences* **275**, 2363-2371.

Rabosky DL, Lovette IJ (2009) Problems detecting density-dependent diversification on phylogenies: reply to Bokma. *Proceedings of the Royal Society B-Biological Sciences* **276**, 995-997.

Rambaut A (2014) *Figtree v.1.4.2*. http://tree.bio.ed.ac.uk/software/figtree/. http://tree.bio.ed.ac.uk/software/figtree/

Rambaut A, Drummond AJ (2013) *Tracer v1.6*. http://tree.bio.ed.ac.uk/software/tracer/. http://tree.bio.ed.ac.uk/software/tracer/

Ramírez SR, Roubik DW, Skov C, Pierce NE (2010) Phylogeny, diversification patterns and historical biogeography of euglossine orchid bees (Hymenoptera: Apidae). *Biological Journal of the Linnean Society* **100**, 552-572.

Ramos ACS, Lemos-Filho JP, Lovato MB (2008) Phylogeographical structure of the Neotropical forest tree *Hymenaea courbaril* (Leguminosae: Caesalpinioideae) and its relationship with the vicariant *Hymenaea stigonocarpa* from Cerrado. *Journal of Heredity* **100**, 206-216.

Ramos ACS, Lemos-Filho JP, Ribeiro RA, Santos FR, Lovato MB (2007) Phylogeography of the tree *Hymenaea stigonocarpa* (Fabaceae: Caesalpinioideae) and the influence of Quaternary climate changes in the Brazilian Cerrado. *Annals of Botany* **100**, 1219-1228.

Rannala B, Yang Z (2013) Improved reversible jump algorithms for Bayesian species delimitation. *Genetics* **194**, 245-253.

Rannala B, Yang Z (2015) Efficient Bayesian species tree inference under the multi-species coalescent. *arXiv preprint arXiv:1512.03843*.

Ratter JA, Ribeiro JF, Bridgewater S (1997) The Brazilian Cerrado vegetation and threats to its biodiversity. *Annals of Botany* **80**, 223-230.

Ree RH, Smith SA (2008) Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic Biology* **57**, 4-14.

Renaud G, Stenzel U, Maricic T, Wiebe V (2015) deML: robust demultiplexing of Illumina sequences using a likelihood-based approach. *Bioinformatics* **31**, 770–772.

Ribeiro J, Colli GR, Caldwell JP, Soares AMVM (2016) An integrated trait-based framework to predict extinction risk and guide conservation planning in biodiversity hotspots. *Biological Conservation* **195**, 214-223.

Ricklefs RE (1987) Community diversity: relative roles of local and regional processes. *Science* **235**, 167-171.

Ricklefs RE (2006) Global variation in the diversification rate of passerine birds. *Ecology* **87**, 2468-2478.

Roelants K, Gower DJ, Wilkinson M, *et al*. (2007) Global patterns of diversification in the history of modern amphibians. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 887-892.

Rohlf FJ, Bookstein FL (1987) A comment on shearing as a method for "size correction". *Systematic Biology*.

Rolland J, Condamine FL, Jiguet F, Morlon H (2014) Faster speciation and reduced extinction in the tropics contribute to the mammalian latitudinal diversity gradient. *Plos Biology* **12**, e1001775.

Ronquist F, Teslenko M, van der Mark P, *et al*. (2012) MrBayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* **61**, 539-542.

Rosenzweig ML (1995) *Species diversity in space and time* Cambridge University Press, Cambridge.

Rull V (2008) Speciation timing and neotropical biodiversity: the Tertiary-Quaternary debate in the light of molecular phylogenetic evidence. *Molecular Ecology* **17**, 2722-2729.

Rundle HD, Nosil P (2005) Ecological speciation. *Ecology Letters* **8**, 336-352.

Saia SEMG, Pessenda LCR, Gouveia SEM, Aravena R, Bendassolli JA (2008) Last glacial maximum (LGM) vegetation changes in the Atlantic Forest, southeastern Brazil. *Quaternary International* **184**, 195-201.

Salerno PE, Señaris JC, Rojas-Runjaic FJM, Cannatella DC (2015) Recent evolutionary history of Lost World endemics: population genetics, species delimitation, and phylogeography of sky-island treefrogs. *Molecular Phylogenetics and Evolution* **82**, 314-323.

Salzburger W, Ewing GB, Von Haeseler A (2011) The performance of phylogenetic algorithms in estimating haplotype genealogies with migration. *Molecular Ecology* **20**, 1952-1963.

San Mauro D, Vences M, Alcobendas M, Zardoya R, Meyer A (2005) Initial diversification of living amphibians predated the breakup of Pangaea. *The American Naturalist* **165**, 590-599.

Santos JC, Coloma LA, Summers K, *et al*. (2009) Amazonian amphibian diversity is primarily derived from late Miocene Andean lineages. *Plos Biology* **7**, 448-461.

Santos MG, Nogueira C, Giugliano LG, Colli GR (2014) Landscape evolution and phylogeography of *Micrablepharus atticolus* (Squamata, Gymnophthalmidae), an endemic lizard of the Brazilian Cerrado. *Journal of Biogeography* **41**, 1506-1519.

Satler JD, Carstens BC, Hedin M (2013) Multilocus species delimitation in a complex of morphologically conserved trapdoor spiders (Mygalomorphae, Antrodiaetidae, Aliatypus). *Systematic Biology* **62**, 805-823.

Scheffers BR, Joppa LN, Pimm SL, Laurance WF (2012) What we know and don't know about Earth's missing biodiversity. *Trends in Ecology & Evolution* **27**, 501-510.

Schwartz MK, McKelvey KS (2008) Why sampling scheme matters: the effect of sampling scheme on landscape genetic results. *Conservation Genetics* **10**, 441-452.

Schweyen H, Rozenberg A, Leese F (2014) Detection and removal of PCR duplicates in population genomic ddRAD studies by addition of a degenerate base region (DBR) in sequencing adapters. *The Biological Bulletin* **227**, 146-160.

Shine R (1979) Sexual selection and sexual dimorphism in the Amphibia. *Copeia* **1979**, 297-306.

Silva JF, Farinas MR, Felfili JM, Klink CA (2006) Spatial heterogeneity, land use and conservation in the cerrado region of Brazil. *Journal of Biogeography* **33**, 536-548.

Silva JMC, Bates JM (2002) Biogeographic patterns and conservation in the South American Cerrado: A tropical savanna hotspot. *Bioscience* **52**, 225-233.

Silva VdN, Pressey RL, Machado RB*, et al*. (2014) Formulating conservation targets for a gap analysis of endemic lizards in a biodiversity hotspot. *Biological Conservation* **180**, 1-10.

Simon MF, Grether R, de Queiroz LP*, et al*. (2009) Recent assembly of the Cerrado, a neotropical plant diversity hotspot, by in situ evolution of adaptations to fire. *Proceedings of the National Academy of Sciences, USA* **106**, 20359-20364.

Simon MF, Pennington T (2012) Evidence for adaptation to fire regimes in the tropical savannas of the Brazilian Cerrado. *International Journal of Plant Sciences* **173**, 711-723.

Sistrom M, Donnellan SC, Hutchinson MN (2013) Delimiting species in recent radiations with low levels of morphological divergence: a case study in Australian *Gehyra* geckos. *Molecular Phylogenetics and Evolution* **68**, 135–143.

Smith BT, McCormack JE, Cuervo AM*, et al*. (2014) The drivers of tropical speciation. *Nature* **515**, 406–409.

Smith SA, Stephens PR, Wiens JJ (2005) Replicate patterns of species richness, historical biogeography, and phylogeny in Holarctic treefrogs. *Evolution* **59**, 2433-2450.

Solís-Lemus C, Knowles LL, Ané C (2015) Bayesian species delimitation combining multiple genes and traits in a unified framework. *Evolution* **69**, 492-507.

Somers KM (1986) Multivariate allometry and removal of size with principal components analysis. *Systematic Biology*.

Stamatakis A (2014) RAxML Version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312-1313.

Stephens M, Donnelly P (2003) A comparison of bayesian methods for haplotype reconstruction from population genotype data. *American journal of human genetics* **73**, 1162-1169.

Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *American journal of human genetics* **68**, 978-989.

Stephens PR, Wiens JJ (2003) Explaining species richness from continents to communities: the time-for-speciation effect in emydid turtles. *The American Naturalist* **161**, 112-128.

Swofford DL (2002) *PAUP* Phylogenetic Analysis using Parsimony (*and other methods). Version 4* Sinauer Associates, Sunderland, Massachusetts.

Tamura K, Peterson D, Peterson N, *et al*. (2011) MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* **28**, 2731-2739.

Team RC (2015) R: a language and environment for statistical computing, Vienna, Austria.

Thomaz AT, Malabarba LR, Bonatto SL, Knowles LL (2015) Testing the effect of palaeodrainages versus habitat stability on genetic divergence in riverine systems: study of a Neotropical fish of the Brazilian coastal Atlantic Forest. *Journal of Biogeography* **42**, 2389–2401.

Thomé MTC, Zamudio KR, Giovanelli JGR, *et al*. (2010) Phylogeography of endemic toads and post-Pliocene persistence of the Brazilian Atlantic Forest. *Molecular Phylogenetics and Evolution* **55**, 1018-1031.

Thornton IWB, Runciman D, Cook S, *et al*. (2002) How important were stepping stones in the colonization of Krakatau? *Biological Journal of the Linnean Society* **77**, 275-317.

Toews DPL, Brelsford A (2012) The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology* **21**, 3907-3930.

Toledo LF, Giovanelli JGR, Giasson LOM, *et al*. (2007) Guia interativo dos anfíbios anuros do Cerrado, Campo rupestre & Pantanal. Editora Neotrópica Ltda, Pinheiros - SP, Brazil.

Udvardy MDF (1975) *A classification of the biogeographical provinces of the world* International Union for Conservation of Nature and Natural Resources, Morges, Switzerland.

Valdujo PH, Silvano DL, Colli GR, Martins M (2012) Anuran species composition and distribution patterns in Brazilian Cerrado, a Neotropical hotspot. *South American Journal of Herpetology* **7**, 63-78.

Van der Hammen T (1974) The Pleistocene changes of vegetation and climate in Tropical South America. *Journal of Biogeography* **1**, 3-26.

Vasconcellos MM, Colli GR, Cannatella D (*in review*) Recurrent dispersal and diversification of treefrogs across distinctive biodiversity hotspots in sub-Amazonian South America. *Evolution*.

Venables WN, Ripley BD (2002) *Modern applied statistics with S* Springer, New York.

Vieites DR, Min M-S, Wake DB (2007) Rapid diversification and dispersal during periods of global warming by plethodontid salamanders. *Proceedings of the National Academy of Sciences* **104**, 19903-19907.

Vuilleumier BS (1971) Pleistocene changes in the fauna and flora of South america. *Science* **173**, 771-780.

Wallace AR (1878) *Tropical nature and other essays* Macmillan, New York.

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358-1370.

Weir JT (2006) Divergent timing and patterns of species accumulation in lowland and highland neotropical birds. *Evolution* **60**, 842-855.

Wells KD (2007) *The Ecology an Behavior of Amphibians* The University of Chicago Press, Chicago, IL.

Werneck FP (2011) The diversification of eastern South American open vegetation biomes: Historical biogeography and perspectives. *Quaternary Science Reviews* **30**, 1630-1648.

Werneck FP, Gamble T, Colli GR, Rodrigues MT, Sites Jr JW (2012a) Deep diversification and long-term persistence in the South American 'dry diagonal': Integrating continent-wide phylogeography and distribution modeling of geckos. *Evolution* **66**, 3014-3034.

Werneck FP, Nogueira C, Colli GR, Sites JW, Costa GC (2012b) Climatic stability in the Brazilian Cerrado: implications for biogeographical connections of South American savannas, species richness and conservation in a biodiversity hotspot. *Journal of Biogeography* **39**, 1695-1706.

Wiens JJ, Donoghue MJ (2004) Historical biogeography, ecology and species richness. *Trends in Ecology & Evolution* **19**, 639-644.

Wiens JJ, Graham CH, Moen DS, Smith SA, Reeder TW (2006) Evolutionary and ecological causes of the latitudinal diversity gradient in hylid frogs: treefrog trees unearth the roots of high tropical diversity. *The American Naturalist* **168**, 579-596.

Wiens JJ, Kuczynski CA, Hua X, Moen DS (2010) An expanded phylogeny of treefrogs (Hylidae) based on nuclear and mitochondrial sequence data. *Molecular Phylogenetics and Evolution* **55**, 871-882.

Wiens JJ, Pyron RA, Moen DS (2011) Phylogenetic origins of local-scale diversity patterns and the causes of Amazonian megadiversity. *Ecology Letters* **14**, 643-652.

Willing E-M, Dreyer C, van Oosterhout C (2012) Estimates of genetic differentiation measured by FST do not necessarily require large sample sizes when using many SNP markers. *Plos One* **7**, e42649.

Willis KJ, Whittaker RJ (2000) Perspectives: paleoecology. The refugial debate. *Science* **287**, 1406-1407.

Woolbright LL (1983) Sexual selection and size dimorphism in anuran Amphibia. *The American Naturalist* **121**, 110-119.

Wright S (1943) Isolation by distance. *Genetics* **28**, 114-138.

Yang Z, Rannala B (2010) Bayesian species delimitation using multilocus sequence data. *PNAS* **107**, 9264-9269.

Yang Z, Rannala B (2014) Unguided species delimitation using DNA sequence data from multiple loci. *Molecular Biology and Evolution* **31**, 3125-3135.

Zachos J (2001) Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* **292**, 686-693.

Zeisset I, Beebee T (2008) Amphibian phylogeography: a model for understanding historical aspects of species distributions. *Heredity* **101**, 109-119.

Zhang C, Zhang DX, Zhu T, Yang Z (2011) Evaluation of a Bayesian coalescent method of species delimitation. *Systematic Biology* **60**, 747–761.

Zhang P, Liang D, Mao RL*, et al*. (2013) Efficient sequencing of anuran mtDNAs and a mitogenomic exploration of the phylogeny and evolution of frogs. *Molecular Biology and Evolution* **30**, 1899-1915.

Zink RM, Barraclough GF (2008) Mitochondrial DNA under siege in avian phylogeography. *Molecular Ecology* **17**, 2107-2121.

Vita

Mariana Mira Vasconcellos was born and raised in Rio de Janeiro, Brazil. After completing her studies in the Military High School in Brasília, she was accepted at Universidade de Brasília from which she graduated with a bachelor's degree in Biology in 2005. Her interests in ecological research started during her first semesters at UnB, being rewarded with two fellowships from CNPq (Brazilian research Institution) that allowed her to conduct undergraduate research for two years. Mariana obtained her Master's degree in Ecology in 2007, working under the supervision of Guarino Colli on population ecology and demography of toads in the Cerrado savanna. She was accepted to the Ecology, Evolution and Behavior program at The University of Texas at Austin to develop her dissertation focusing on the evolution of treefrogs in the Cerrado savanna. She was awarded a Fulbright fellowship in partnership with CAPES (research foundation in Brazil) to develop her doctoral studies in the US, and an NSF Doctoral Dissertation Improvement Grant that funded part of her doctoral research. Next, Mariana will be joining the Carnaval lab at CUNY in New York for two years before returning to her home country, where she will pursue a Professorship at a high-profile University or research Institution.

Address (or e-mail): marianamv@utexas.edu

This dissertation was typed by Mariana M. Vasconcellos.